USGS

science for a changing world

A Practical Primer on Geostatistics

By Ricardo A. Olea

U.S. Department of the Interior
U.S. Geological Survey

For product and ordering information:
World Wide Web: http://www.usgs.gov/pubprod
Telephone: 1-888-ASK-USGS (1-888-275-8747)

For more information on the USGS—the Federal source for science about the Earth,
its natural and living resources, natural hazards, and the environment:
World Wide Web: http://www.usgs.gov
Telephone: 1-888-ASK-USGS (1-888-275-8747)

USGS

science for a changing world

A Practical Primer on Geostatistics

By Ricardo A. Olea

Open-File Report 2009-1103

CONTENTS

1. INTRODUCTION

THE CHALLENGE

- Most geological phenomena are extraordinarily complex in their interrelationships and vast in their geographical extension.

- Ordinarily, engineers and geoscientists are faced with corporate or scientific requirements to properly prepare geological models with measurements involving a small fraction of the entire area or volume of interest.

- Exact description of a system such as an oil reservoir is neither feasible nor economically possible.

- The results are necessarily uncertain.

  Note that the uncertainty is not an intrinsic property of the systems; it is the result of incomplete knowledge by the observer.

THE AIM OF GEOSTATISTICS

- The main objective of geostatistics is the characterization of spatial systems that are incompletely known, systems that are common in geology.

- A key difference from classical statistics is that geostatistics uses the sampling location of every measurement.

Two-dimensional example

Unless the measurements show spatial correlation, the application of geostatistics is pointless.

Ordinarily the need for additional knowledge goes beyond a few points, which explains the display of results graphically as fishnet plots, block diagrams, and maps.
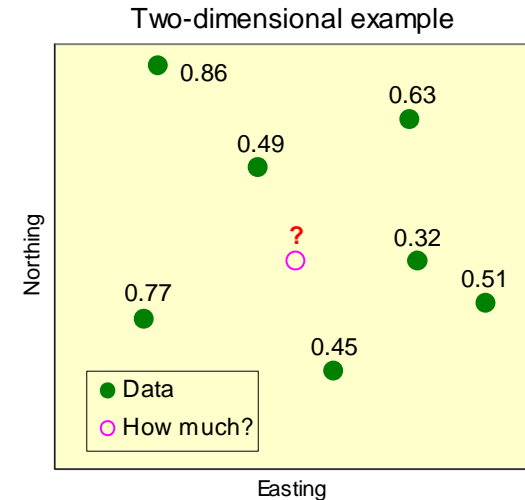
GEOSTATISTICAL METHODS

Geostatistics is a collection of numerical techniques for the characterization of spatial attributes using primarily two tools:

- probabilistic models, which are used for spatial data in a manner similar to the way in which time-series analysis characterizes temporal data, or
- pattern recognition techniques.

The probabilistic models are used as a way to handle uncertainty in results away from sampling locations, making a radical departure from alternative approaches like inverse distance estimation methods.

DIFFERENCES WITH TIME SERIES

On dealing with time-series analysis, users frequently concentrate their attention on extrapolations for making forecasts.
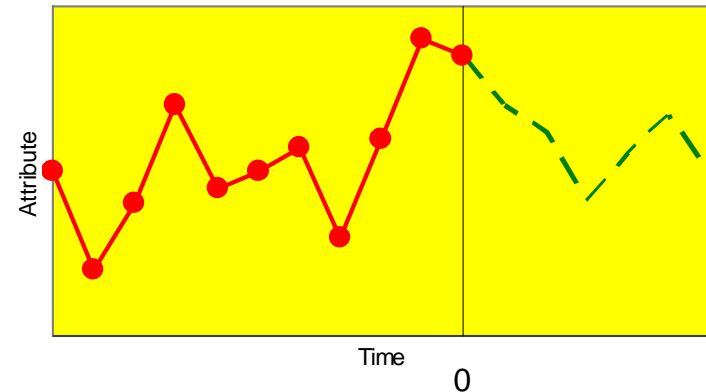
Although users of geostatistics may be interested in extrapolation, the methods work at their best interpolating.
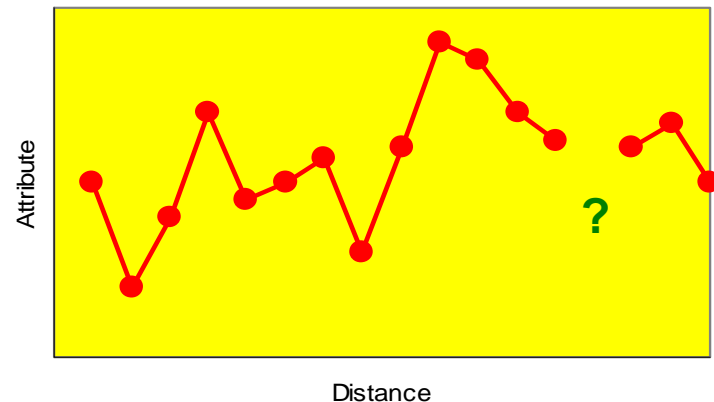
This simple difference has significant methodological implications.

HISTORICAL REMARKS

- As a discipline, geostatistics was firmly established in the 1960s by the French engineer Georges Matheron, who was interested in the appraisal of ore reserves in mining.

- Geostatistics did not develop overnight. Like other disciplines, it has built on previous results, many of which were formulated with different objectives in various fields.

PIONEERS

Seminal ideas conceptually related to what today we call geostatistics or spatial statistics are found in the work of several pioneers, including:

- 1940s: A.N. Kolmogorov in turbulent flow and N. Wiener in stochastic processing
- 1950s: D. Krige in mining
- 1960s: B. Mathern in forestry and L.S. Gandin in meteorology

USGS

CALCULATIONS

Serious applications of geostatistics require the use of digital computers.

Although for most geostatistical techniques rudimentary implementation from scratch is fairly straightforward, coding programs from scratch is recommended only as part of a practice that may help users to gain a better grasp of the formulations.

SOFTWARE

For professional work, the reader should employ software packages that have been thoroughly tested to handle any sampling scheme, that run as efficiently as possible, and that offer graphic capabilities for the analysis and display of results.

This primer employs primarily the package Stanford Geomodeling Software (SGeMS)—recently developed at the Energy Resources Engineering Department at Stanford University—as a way to show how to obtain results practically.

This applied side of the primer should not be interpreted as the notes being a manual for the use of SGeMS. The main objective of the primer is to help the reader gain an understanding of the fundamental concepts and tools in geostatistics.

ORGANIZATION OF THE PRIMER

The chapters of greatest importance are those covering kriging and simulation. All other materials are peripheral and are included for better comprehension of these main geostatistical modeling tools.

The choice of kriging versus simulation is often a big puzzle to the uninitiated, let alone the different variants of both of them. Chapters 14, 18, and 19 are intended to shed light on those subjects.

The critical aspect of assessing and modeling spatial correlation is covered in chapter 7.

Chapters 2 and 3 review relevant concepts in classical statistics.

COURSE OBJECTIVES

This course offers stochastic solutions to common problems in the characterization of complex geological systems.

At the end of the course, participants should have:

- an understanding of the theoretical foundations of geostatistics;
- a good grasp of its possibilities and limitations; and
- reasonable familiarity with the SGeMS software, thus opening the possibility of practically applying geostatistics.

2. UNIVARIATE STATISTICS

EVERYTHING AND A PIECE

In statistics, a population is the collection of all possible outcomes or individuals comprising the complete system of interest; for example, all people in the Unites States.

Populations may be hard or impossible to analyze exhaustively. In statistics, a limited collection of measurements is called a sample; for example, a Gallup Poll.

Unfortunately, the term "sample" is employed with different meanings in geology and statistics.

Geology	Statistics
collection	sample
sample	observation

The statistical usage of the term sample is observed in what follows.

RANDOM VARIABLE
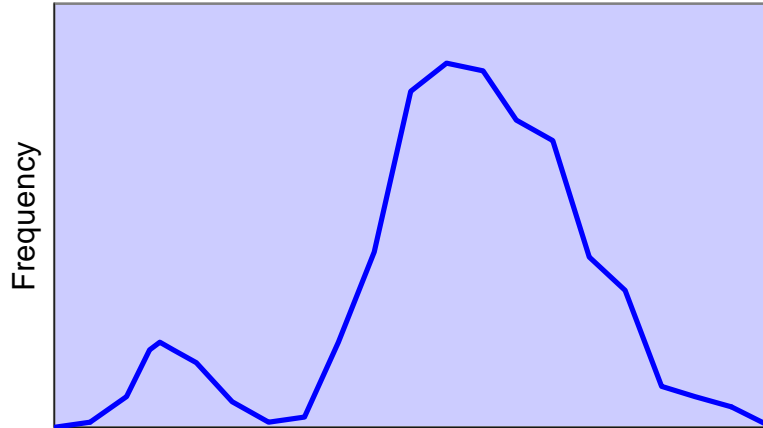
A random variable or variate is a quantity that may take any of the values within a given set with specified relative frequencies.

The concept is used in geostatistics extensively to characterize a population or convey the unknown value that an attribute may take at any spatiotemporal location.

DESCRIPTIVE ANALYSIS

- A sample of an attribute ordinarily comprises several measurements, which are best understood when organized in some way. This is an important aspect of statistics.

- The number of measurements in a sample is the sample size.

- There are multiple options to make the data more intelligible. Some of these are more convenient than others, depending on factors such as the sample size and the ultimate objectives of the study.

SAMPLE VISUALIZATION

FREQUENCY TABLE

Given some numerical information, if the interval of variation of the data is divided into class intervals— customarily of the same lengths—and all observations are assigned to their corresponding classes, the result is a count of relative frequency of the classes.

USGS

UNCF EXAMPLE FREQUENCY TABLE

	Class	Count	Frequency, %
1	7,680-7,710	1	1.43
2	7,710-7,740	1	1.43
3	7,740-7,770	1	1.43
4	7,770-7,800	2	2.86
5	7,800-7,830	5	7.14
6	7,830-7,860	6	8.57
7	7,860-7,890	10	14.29
8	7,890-7,920	11	15.71
9	7,920-7,950	13	18.57
10	7,950-7,980	7	10.00
11	7,980-8,010	7	10.00
12	8,010-8,040	4	5.71
13	8,040-8,070	2	1.43
Total		70	100.00

This example from a major oil company relates to depth in feet to an unconformity (UNCF) in an undisclosed area. It will be used as a common reference to graphically illustrate other definitions in this chapter.
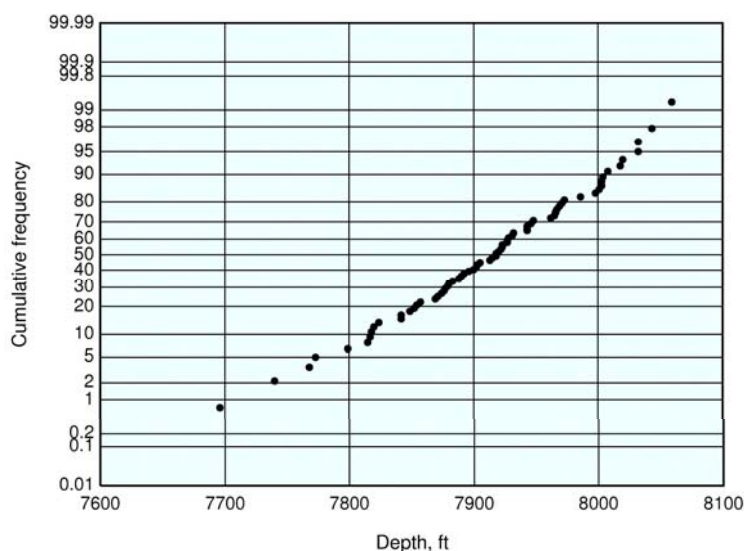
HISTOGRAM
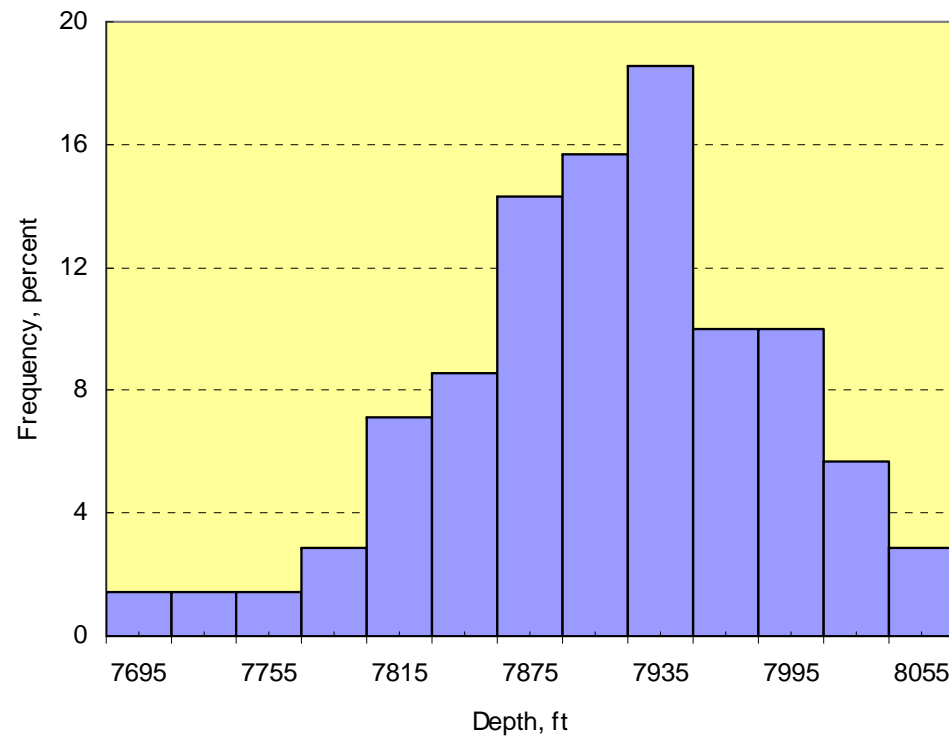
A histogram is a graphical representation of a frequency table.

CUMULATIVE FREQUENCY

Summaries based on frequency tables depend on the selection of the class interval and origin.

Given a sample of size n, this drawback is eliminated by displaying each observation z_i versus the proportion of the sample that is not larger than z_i.

Each proportion is a multiple of $100/n$. The vertical axis is divided in n intervals and the data are displayed at the center of the corresponding interval.

Customarily, the vertical axis is scaled so that data from a normal distribution (page 49) display as a straight line.

SUMMARY STATISTICS

SUMMARY STATISTIC

Summary statistics are complementary or alternative descriptors to histograms and cumulative distributions.

A statistic is a synoptic value that is calculated from a sample of observations, which is usually, but not necessarily, an estimator of some population parameter.

Generally, summary statistics are subdivided into three categories:

- Measures of location or centrality
- Measures of spread or dispersion
- Measures of shape

MEASURES OF LOCATION

Measures of location give an idea about the central tendency of the data. They are:

- mean

- median

- mode

MEAN

The arithmetic mean or simply the mean, $\hat{m}$, of a sample of size n is the additive average of all the observations, z_i:

$$\hat{m} = \frac{1}{n} \sum_{i=1}^{n} z_i.$$
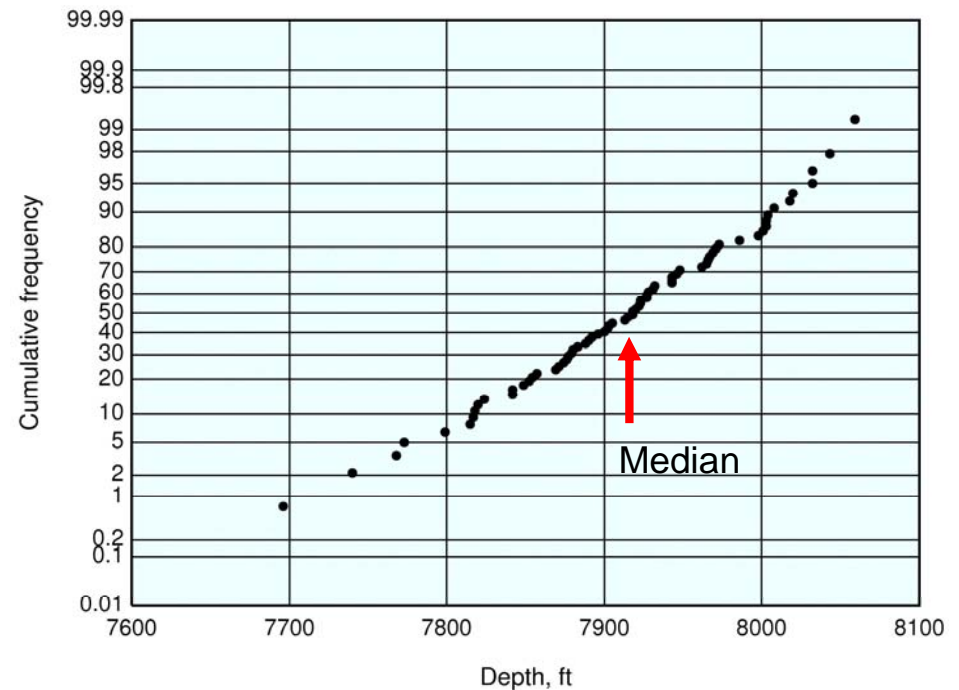
The mean of the UNCF sample is 7,912.2 ft.
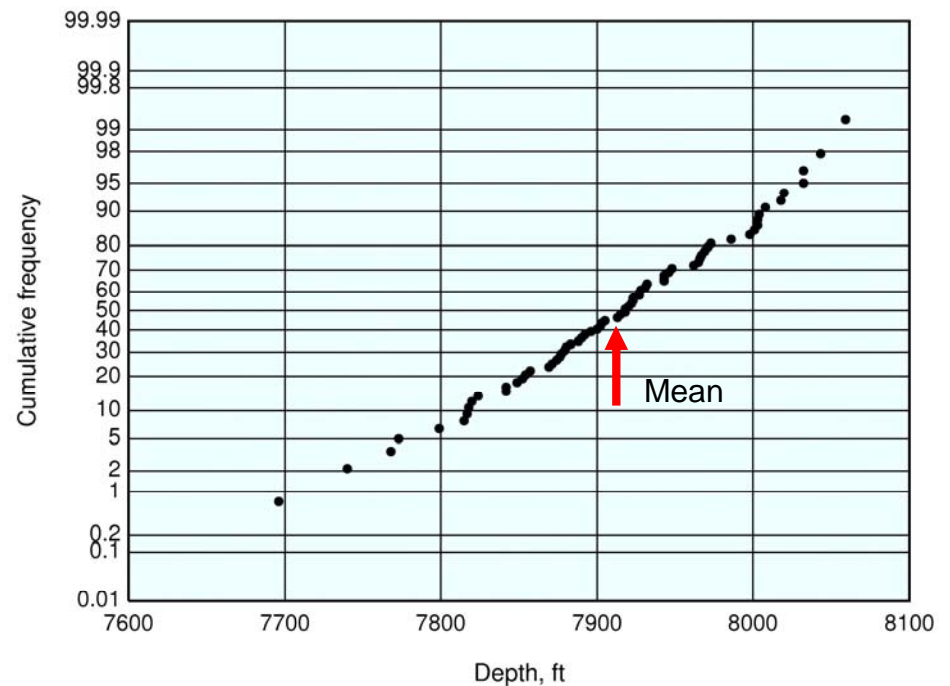
MEDIAN

The median, Q_2, of a sample is the value that evenly splits the number of observations z_i into a lower half of smaller observations and an upper half of larger measurements.
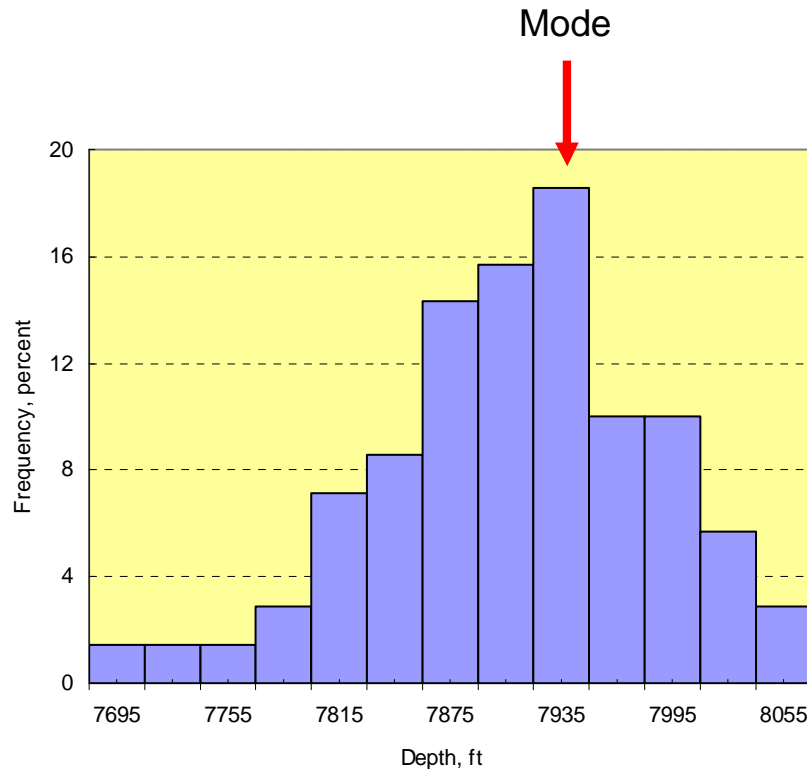
If z_i is sorted by increasing values, then

$$Q_2 = \begin{cases} z_{(n+1)/2}, & \text{if } n \text{ is odd,} \\ 0.5 \cdot \left(z_{n/2} + z_{(n/2)+1} \right), & \text{if } n \text{ is even.} \end{cases}$$

The median of the UNCF sample is 7,918 ft.

MODE

The mode of a sample is the most probable or frequent value, or, equivalently, the center point of the class containing the most observations.

For the UNCF sample, the center of the class with the most observations is 7,935 ft.

ROBUSTNESS

Robustness denotes the ability of statistical methods to work well not only under ideal conditions but also in the presence of data problems, mild to moderate departures from assumptions, or both.

For example, in the presence of large errors, the median is a more robust statistic than the mean.

MEASURES OF SPREAD

Measures of spread provide an idea of the dispersion of the data. The most common measures are:

- variance
- standard deviation
- extreme values
- quantiles
- interquartile range

VARIANCE

The variance, $\hat{\sigma}^2$, is the average squared dispersion around the mean:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - m)^2 = \frac{1}{n}\left[\sum_{i=1}^{n} z_i^{\,2} - n \cdot m^2\right],$$

expressions that are commonly restricted to estimate variances of finite populations.

When dealing with samples, the denominator is often changed to n - 1.

Because this is a quadratic measure, it is less robust than most other measures of spread.

The variance of the UNCF sample is 5,474 sq ft.

STANDARD DEVIATION

The standard deviation is the positive square root of the variance.

It has the advantage of being in the same units as the attribute.

The standard deviation of the UNCF sample is 74.5 ft.

According to Chebyshev's theorem, for any sample and $t > 1$, the proportion of data that deviates from the mean $\hat{m}$ at least $t \cdot \hat{\sigma}$ is at most t^{-2}:

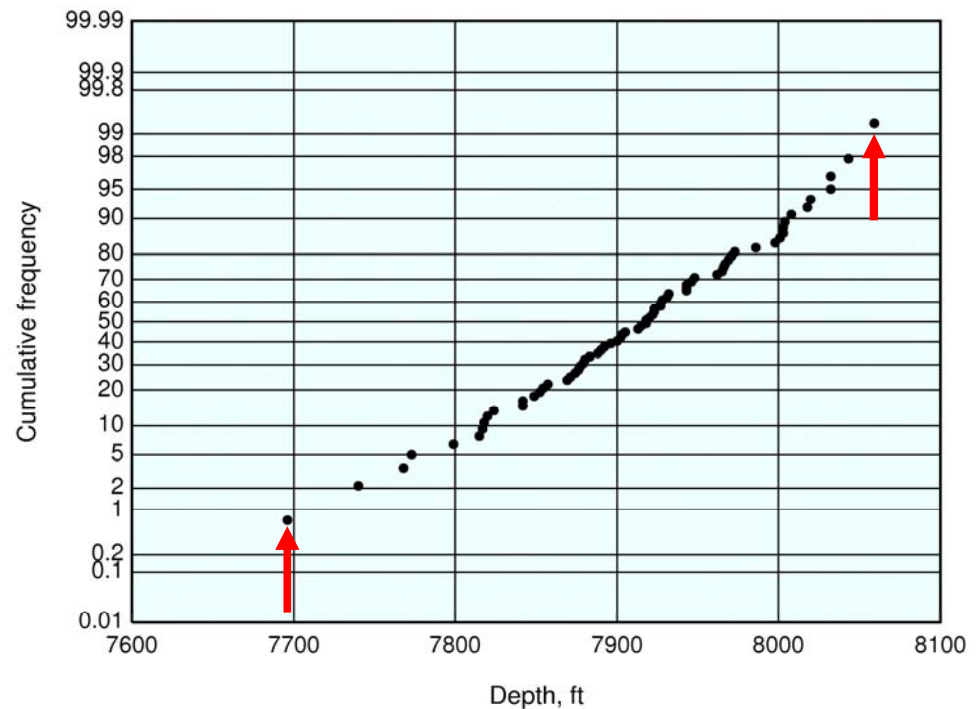$$\text{Prop}\left(\left|X - \hat{m}\right| \geq t \cdot \hat{\sigma}\right) \leq \frac{1}{t^2}$$
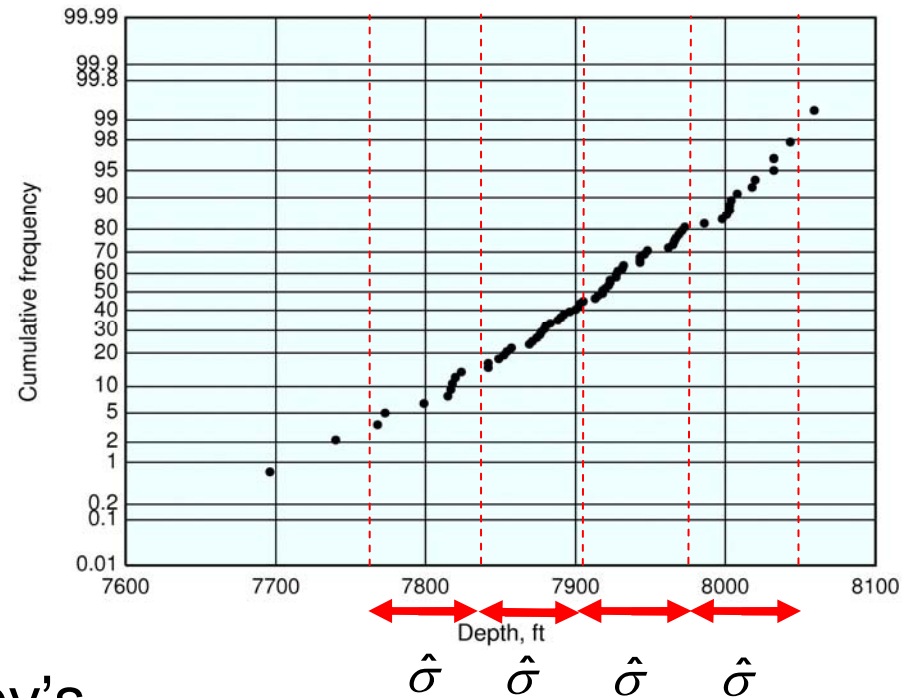
EXTREME VALUES

The extreme values are the minimum and the maximum.

For the UNCF sample, the minimum value is 7,696 ft and the maximum value is 8,059 ft.

This measure is not particularly robust, especially for small samples.

QUANTILES

The idea of the median splitting the ranked sample into two equal-size halves can be generalized to any number of partitions with equal numbers of observations. The partition boundaries are called quantiles or fractiles. The names for the most common quantiles are:

- Median, for 2 partitions
- Quartiles, for 4 partitions
- Deciles, for 10 partitions
- Percentiles, for 100 partitions

The number of boundaries is always one less than the number of partitions.

UNCF QUARTILES

$Q_1 = 7,871.75$ ft

$Q_2 = 7,918$ ft

$Q_3 = 7,965.75$ ft
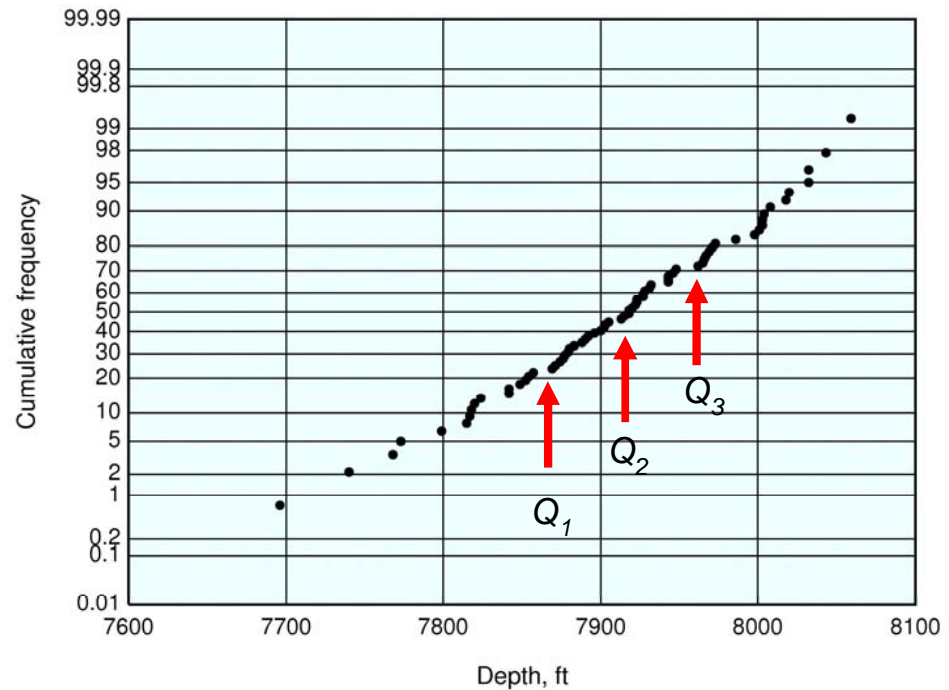
Q_2 coincides with the median.

INTERQUARTILE RANGE

The interquartile range, *iqr,* is the difference between the upper and the lower quartiles

$$iqr = Q_3 - Q_1,$$

thus measuring the central spread of the data.

For the UNCF sample, the interquartile range is 94 ft.

The interquantile range is more robust than the variance but insensitive to values in the lower and upper tails.
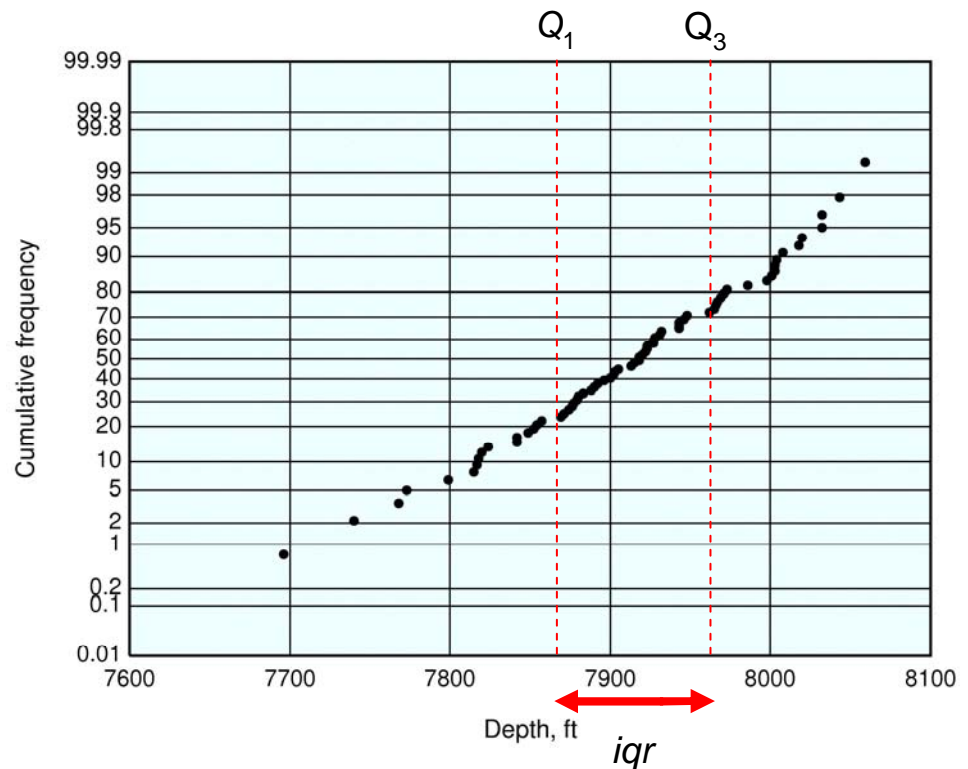
OUTLIER

Outliers are values so markedly different from the rest of the sample that they rise the suspicion that they may be from a different population or that they may be in error, doubts that frequently are hard to clarify. In any sample, outliers are always few, if any.

A practical rule of thumb is to regard any value deviating more than 1.5 times the interquartile range, *iqr*, from the median as a mild outlier and a value departing more than 3 times *iqr* as an extreme outlier.
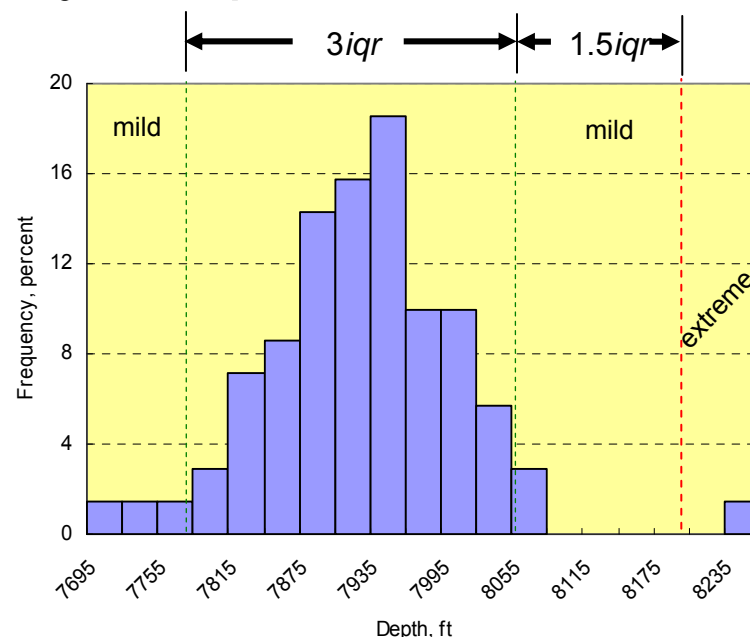

For the UNCF sample, all mild outliers seem to be legitimate values, while the extreme outlier of 8,240 ft is an error.

BOX-AND-WHISKER PLOT

The box-and whisker plot is a
simple graphical way to summarize
several of the statistics:

- Minimum
- Quartiles
- Maximum
- Mean

Variations in this presentation
style abound. Extremes may
exclude outliers, in which case
the outliers are individually posted
as open circles. Extremes
sometimes are replaced by the 5
and 95 percentiles.

UNCF

Depth, ft

8100
8000
7900
7800
7700
7600

Max.

Q_3
Q_2
Mean
Q_1

Min.
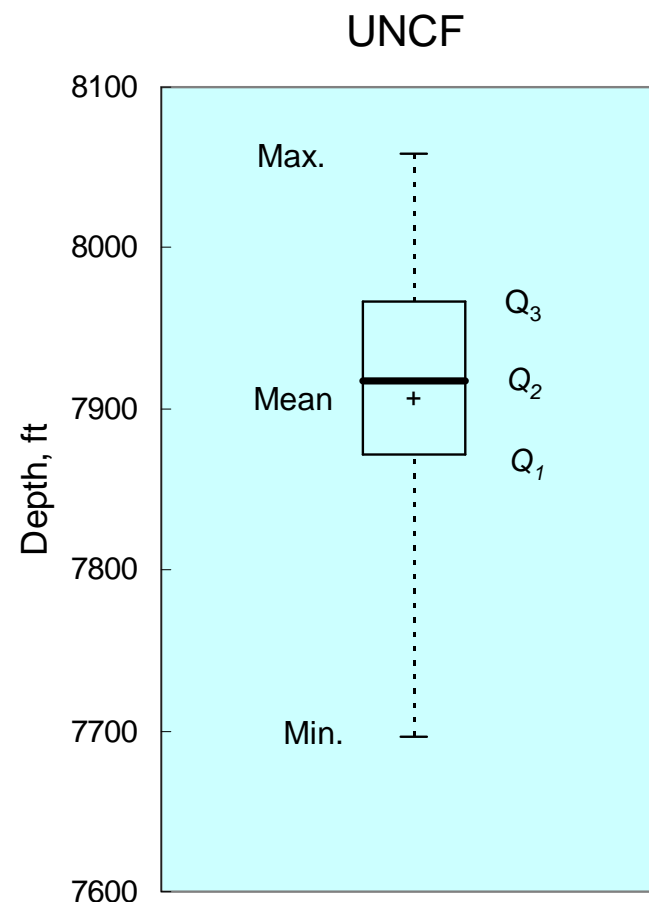
MEASURES OF SHAPE

The most commonly used measures of shape in the distribution of values are:

- Coefficient of skewness
- Quartile skew coefficient
- Coefficient of kurtosis

COEFFICIENT OF SKEWNESS

The coefficient of skewness is a measure of asymmetry of the histogram. It is given by:

$$B_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - m)^3}{\sigma^3}$$

If : $B_1 < 0$, the left tail is
$\qquad$ longer;

$\quad B_1 = 0$, the distribution is
$\qquad$ symmetric;

$\quad B_1 > 0$, the right tail is
$\qquad$ longer.

The UNCF coefficient of skewness is -0.38.

$B_1 = -0.38$

Frequency, percent

Depth, ft
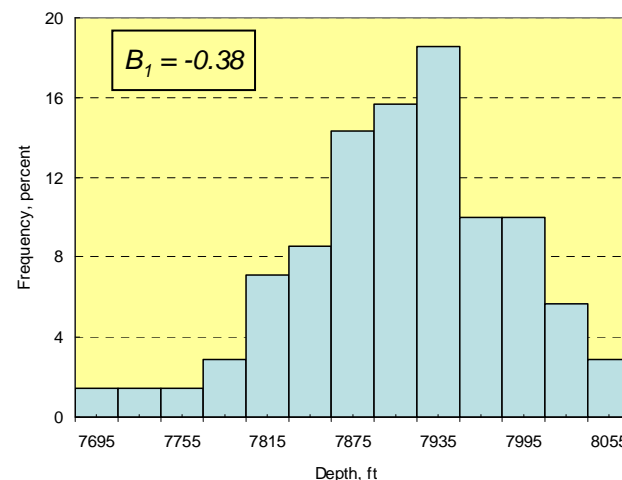
QUARTILE SKEW COEFFICIENT

The quartile skew coefficient serves the same purpose as the coefficient of skewness, but it is more robust, yet only sensitive to the central part of the distribution. Its definition is:

$$qs = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{iqr}$$

If : $qs < 0$, the left tail is longer;

$qs = 0$, the distribution is symmetric;

$qs > 0$, the right tail is longer.

The UNCF quartile skew coefficient is 0.02.

41
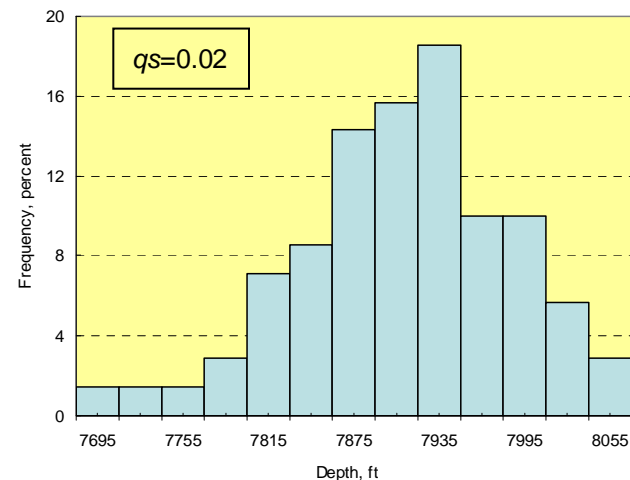
COEFFICIENT OF KURTOSIS

This statistic measures the concentration of values around the mean. Its definition is:

$$B_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - m)^4}{\sigma^4}$$

If : $B_2 < 3$, the distribution is more peaked than the Gaussian distribution;

$B_2 = 3$, it is as peaked as the Gaussian;

$B_2 > 3$, it is less peaked than the Gaussian.

The UNCF coefficient of kurtosis is 3.08.

Gaussian distribution
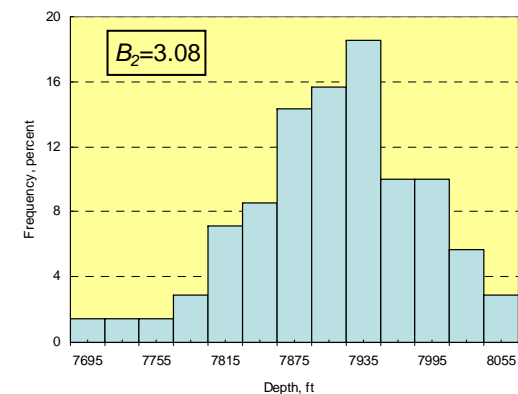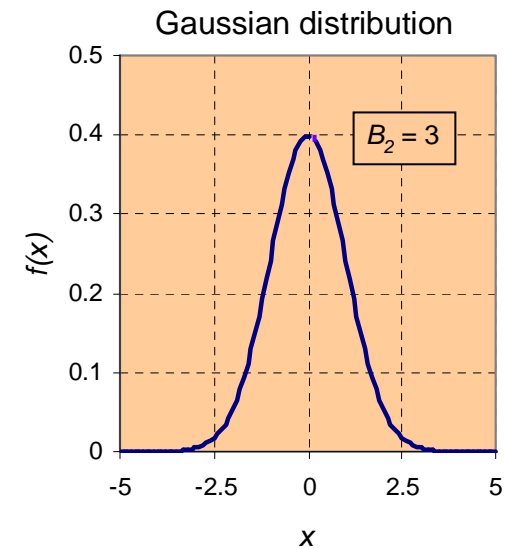
MODELS

PROBABILITY

Probability is a measure of the likelihood that an event, *A*, may occur. It is commonly denoted by Pr[*A*].

- The probability of an impossible event is zero, Pr[*A*] = 0. It is the lowest possible probability.

- The maximum probability is Pr[*A*] = 1, which denotes certainty.

- When two events *A* and *B* cannot take place simultaneously, Pr[*A* or *B*] = Pr[*A*] + Pr[*B*].

- Frequentists claim that Pr[*A*] = N_A/N, where *N* is total number of outcomes and N_A the number of outcomes of *A*. The outcomes can be counted theoretically or experimentally.

- For others, a probability is a degree of belief in *A*, even if no random process is involved nor a count is possible.

CONDITIONAL PROBABILITY

Let *A* and *B* be two random events. The conditional probability of *A* given *B*, Pr[*A* | *B*], is the probability of *A* for the special case when *B* has occurred. If

$$\Pr[A \mid B] \neq \Pr[A],$$

then the result of one affects the probability of the other. Such type of outcomes are said to be statistically dependent. Otherwise, they are statistically independent. For example:

- Head and tail in successive flips of a fair coin are independent events.

- Being dealt a king from a deck of cards and having two kings on the table are dependent events.

BAYES'S THEOREM

This is one of the most widely used probability relationships for the calculation of conditional probabilities. Given outcomes *A* and *B*:

$$\Pr[A\,|\,B] = \frac{\Pr[B\,|\,A]}{\Pr[B]}\Pr[A]$$

Example
Suppose there are two boxes. A blue ball is drawn (event *B*). What is the probability the ball came from box #1 (event *A*)?

	Number of balls		
Box	Blue	Red	Total
#1	20	5	25
#2	12	18	30
	32	23	55

- If one only knows that there are 2 boxes, Pr[*A*] = ½ = 0.5.

- Now, if the table is available, Pr[*B*|*A*] = 20/25 = 0.8.

- Pr[*B*] = 32/55 = 0.59. Hence:

$$\Pr[A\,|\,B] = \frac{0.8}{0.59}\,0.5 = 0.69$$

PROBABILITY FUNCTIONS

Analytical functions approximating experimental fluctuations are the alternative to numerical descriptors and measures. They provide approximations of general conditions. Their drawback is the loss of fine detail in favor of simpler models.

Models approximating histograms are called probability density functions.

Variations in the parameters of a probability density function allow the generation of a family of distributions, sometimes with radically different shapes.

The main subdivision is into discrete and continuous distributions, of which the binomial and normal distribution, respectively, are typical and common examples.

BINOMIAL DISTRIBUTION

This is the discrete probability density function, $f(x; p, n)$, of the number of successes in a series of independent (Bernoulli) trials, such as head or tails in coin flipping. If the probability of success at every trial is p, the probability of x successes in n independent trials is

$$f(x; p, n) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \ x = 0, 1, 2, \ldots, n$$

USGS

48
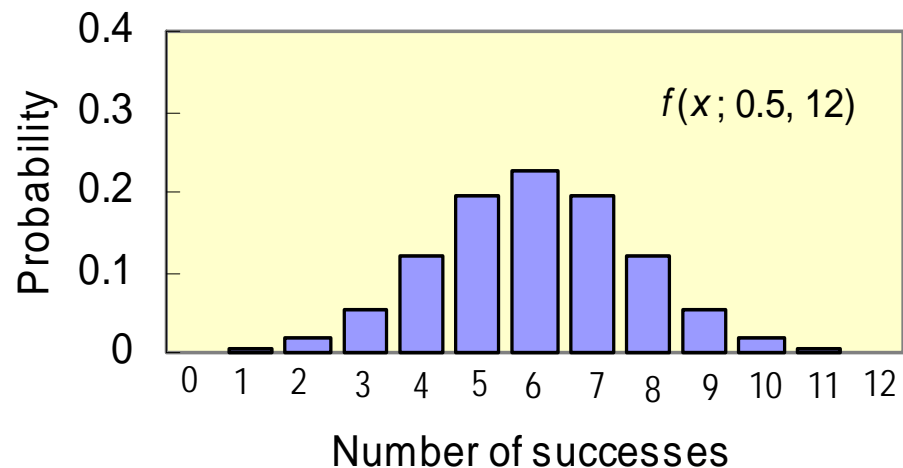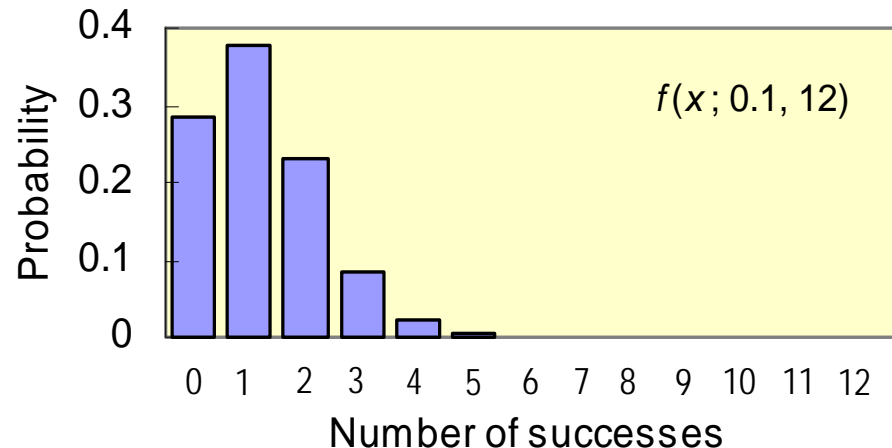
NORMAL DISTRIBUTION
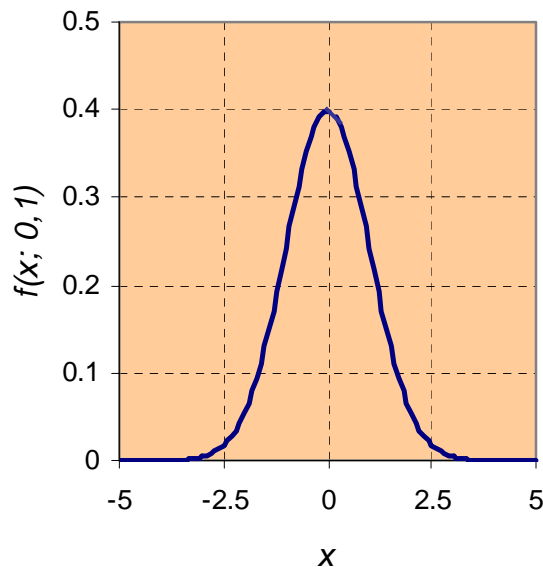
The most versatile of all continuous models is the normal distribution, also known as the Gaussian distribution. Its parameters are μ and σ, which coincide with the mean and the standard deviation.

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$

If $X = \log(Y)$ is normally distributed, Y is said to follow a lognormal distribution. Lognormal distributions are positively defined and positively skewed.

PROBABILITY FROM MODELS

$$\text{Prob}[X \le x_1] = \int_{-\infty}^{x_1} f(x)dx \qquad\qquad \text{Prob}[X \le x_1] = F(x_1)$$

Examples:

Gaussian distribution

Gaussian distribution
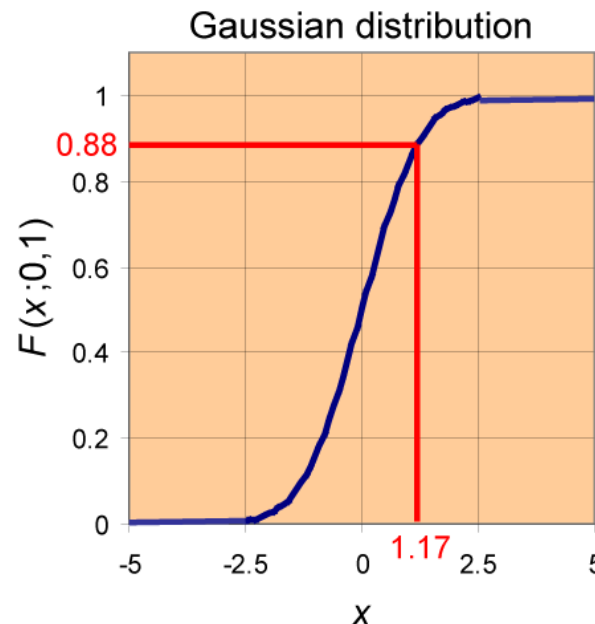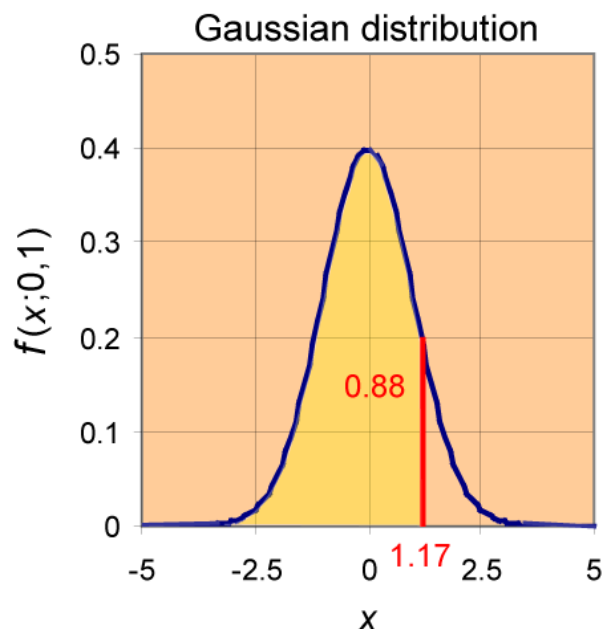
$$\text{Prob}[X \le 1.17] = \int_{-\infty}^{1.17} \text{Normal}(x;0,1)dx$$
$$= 0.88$$

$$\text{Prob}[X \le 1.17] = F(1.17)$$
$$= 0.88$$

50

EXPECTED VALUE

Let X be a random variable having a probability distribution $f(x)$ and let $u(x)$ be a function of x. The expected value of $u(x)$ is denoted by the operator $\mathrm{E}[u(x)]$ and it is the probability weighted average value of $u(x)$.

If X is continuous, such as temperature:

$$\mathrm{E}[u(x)] = \int_{-\infty}^{\infty} u(x)f(x)dx,$$

and if it is discrete, like in coin flipping:

$$\mathrm{E}[u(x)] = \sum_{x} u(x)f(x).$$

In the latter case, for the trivial example of $u(x) = x$, if all values are equally probable, the expected value turns into

$$\mathrm{E}[x] = \frac{1}{n}\sum_{x} x,$$

which is exactly the definition of the mean.

MOMENT

Moment is the name given to the expected value when the function of the random variable, if it exists, takes the form $(x-a)^k$, where k is an integer larger than zero, called the order.

If a is the mean, then the moment is a central moment.

The central moment of order 2 is the variance. For an equally probable discrete case,

$$M_2 = \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$$

3. BIVARIATE STATISTICS

TOOLS

Frequently there is interest in comparing two or more measurements made for the same object or site. Among the most common classical alternatives, we have:

- Scatterplot
- Correlation coefficient
- Regression
- Quantile-quantile plot
- Probability-probability plot

Some of these concepts can be generalized to deal with more than two variables.

SCATTERPLOT

A bivariate scatterplot is a Cartesian posting in which the abscissa and the ordinate are any two variables consistently measured for a series of objects.

 Scatterplots are prepared for exploring or revealing form, direction, and strength of association between two attributes.

Mecklenburg Bay sea floor, Germany
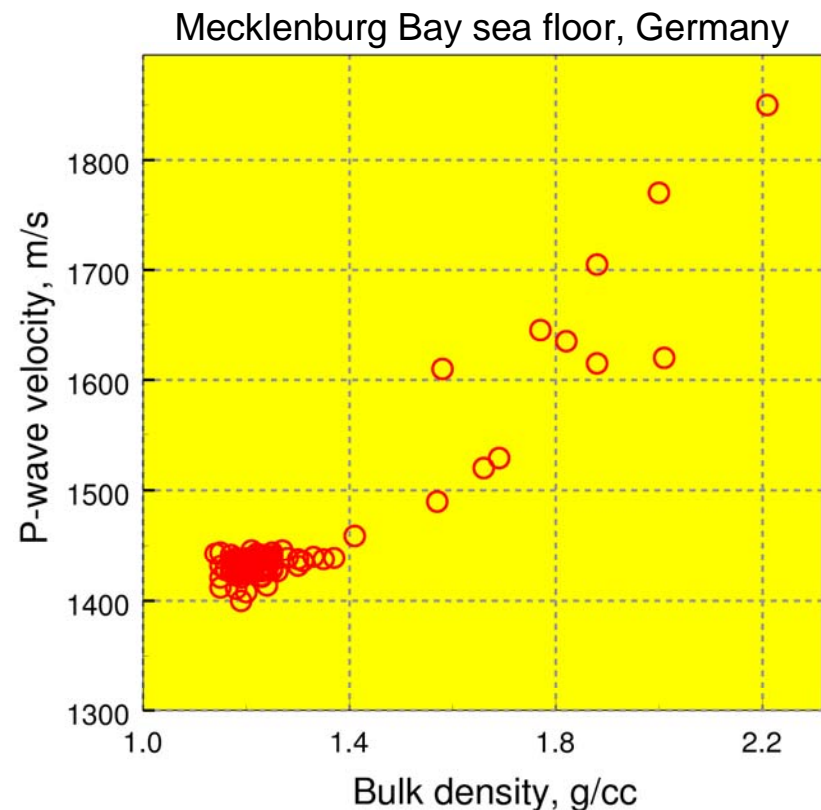
COVARIANCE

The covariance is a measure of joint variation.

Given two random variables X and Y with means μ_x and μ_Y, their covariance is the expected value:

$$\text{Cov}_{X,Y} = E\big[(X - \mu_X)(Y - \mu_Y)\big]$$

The covariance estimator when using a sample of point measurements is:

$$\hat{\text{Cov}}_{X,Y} = \frac{1}{n-1}\sum_{i=1}^{n} x_i \cdot y_i - \frac{1}{n \cdot (n-1)}\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i.$$

CORRELATION COEFFICIENT

This coefficient is the number most commonly used to summarize bivariate comparisons. If σ_X and σ_Y are the standard deviations for two variables, their correlation coefficient, ρ, is given by:

$$\rho = \frac{\text{Cov}_{X,Y}}{\sigma_X \cdot \sigma_Y}$$

- It only makes sense to employ ρ for assessing linear associations.
- ρ varies continuously from -1 to 1:
   - 1, perfect direct linear correlation
   - 0, no linear correlation
   - -1, perfectly inverse correlation

1

0

-0.5

-1
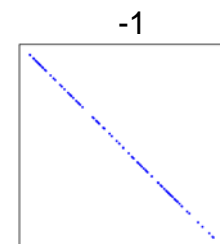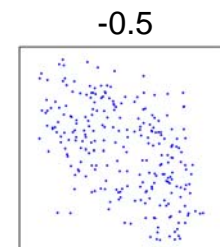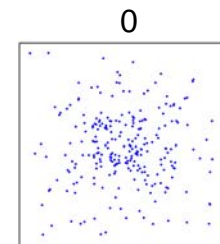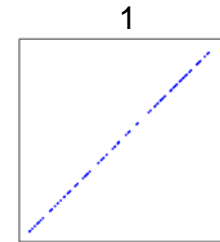
REGRESSION

Regression is a method for establishing analytical dependency of one or more variables on another mainly to determine the degree of their dependency, to estimate values not included in the sample, or to summarize the sample.

- The variable in the abscissa is called the regressor, independent, or explanatory variable.

- The variable in the ordinate is the regressed, dependent, or response variable.

- In many studies, which variable goes into which axis is an arbitrary decision, but the result is different. Causality or the physics of the process may help in resolving the indetermination.

REGRESSION MODEL

Response variable (y)

$f(x_i; \mathbf{\theta})$

ε_i

Explanatory variable (x)
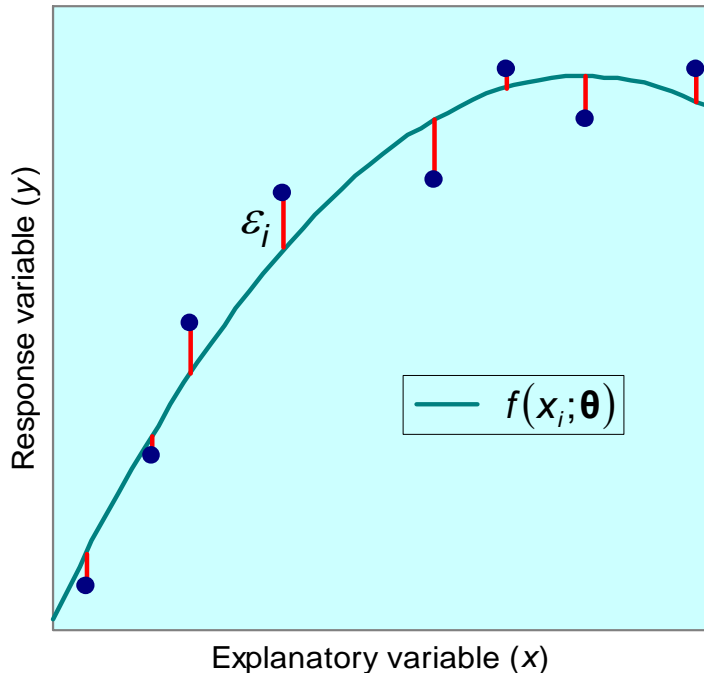
- The model is:
$$y_i = f(x_i; \mathbf{\theta}) + \varepsilon_i$$

- $f(x_i; \mathbf{\theta})$ is any continuous function of x that is judiciously selected by the user. $\mathbf{\theta}$ are unknown parameters.

- Term ε is a random variable accounting for the error.

- Having selected $f(x_i; \mathbf{\theta})$, parameters $\mathbf{\theta}$ are calculated by minimizing total error, for which there are several methods.

Avoid applying the model outside the extreme values of the explanatory variable.

LINEAR REGRESSION

The simplest regression case is one in which the parameters are estimated by minimizing the mean square error, r^2, associated with a linear model,

$$r^2 = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2$$

In this special situation, ρ^2 accounts for the proportion of variation accounted for by the regression.

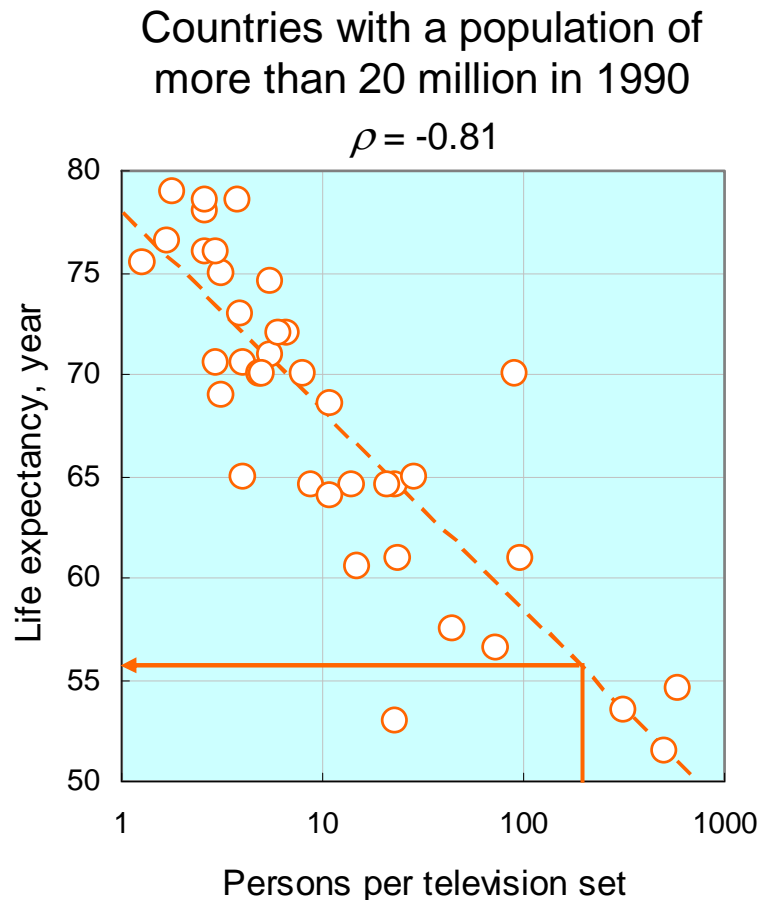In the example, $\rho = 0.94$. Hence, in this case, the linear regression explains 88% $\left(100 \cdot \rho^2\right)$ of the variation.

Mecklenburg Bay seafloor, Germany

$\rho = 0.94$

P-wave velocity, m/s

Bulk density, g/cc

NONLINEAR REGRESSION

Linear

$r^2 = 25.47$

Quadratic polynomial

$r^2 = 20.56$

Cubic polynomial

$r^2 = 20.33$

Sixth degree polynomial

$r^2 = 19.98$
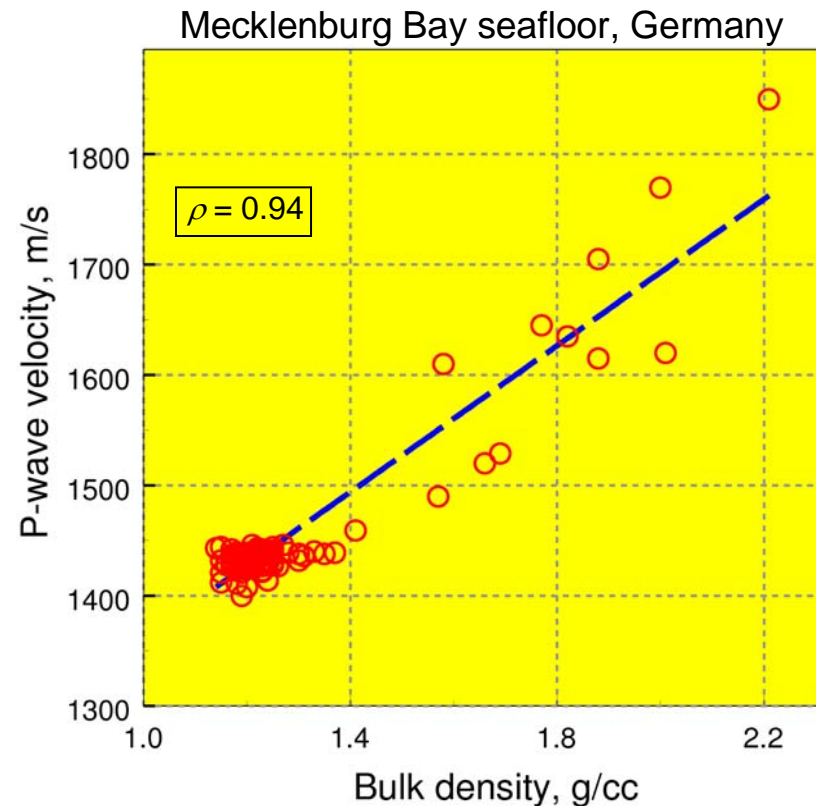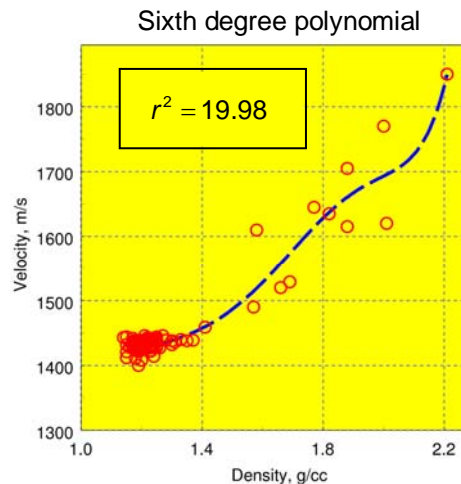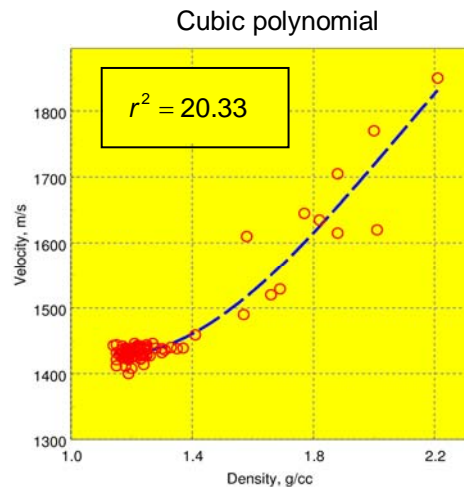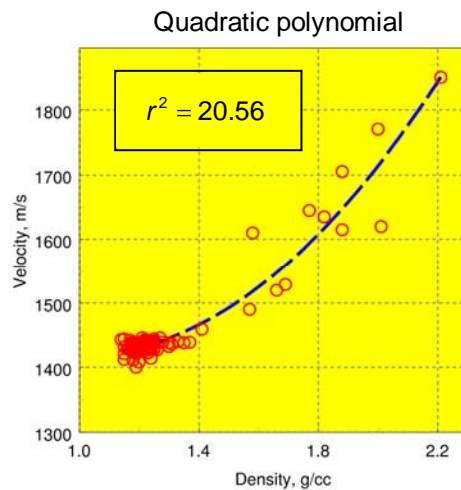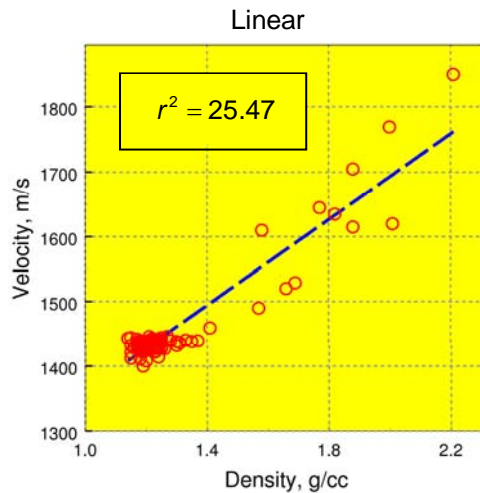
- In theory, the higher the polynomial degree, the better the fit.

- In practice, the higher the polynomial, the less robust the solution.

- Overfitting may capture noise and not systematic variation.

IMPLICATIONS

Countries with a population of
more than 20 million in 1990

$\rho = -0.81$

Life expectancy, year (y-axis)
Persons per television set (x-axis)

High to good correlation:
- allows prediction of one variable when only the other is known, but the inference may be inaccurate, particularly if the correlation coefficient is low;
- means the variables are related, but the association may be caused by a common link to a third lurking variable, making the relationship meaningless;
- does not necessarily imply cause and effect.

QUANTILE-QUANTILE PLOT

- A quantile-quantile or Q-Q plot is a scatterplot based on ranked data.

- The pairing is independent of the object or site where the observations were taken. The first pair of coordinates has the minimum value for each attribute, the second pair is made of the second smallest readings, and so on until finishing with the two maximum values. Interpolations are necessary for different size samples.

- Q-Q plots are sensitive to a shift and scaling of the distributions.
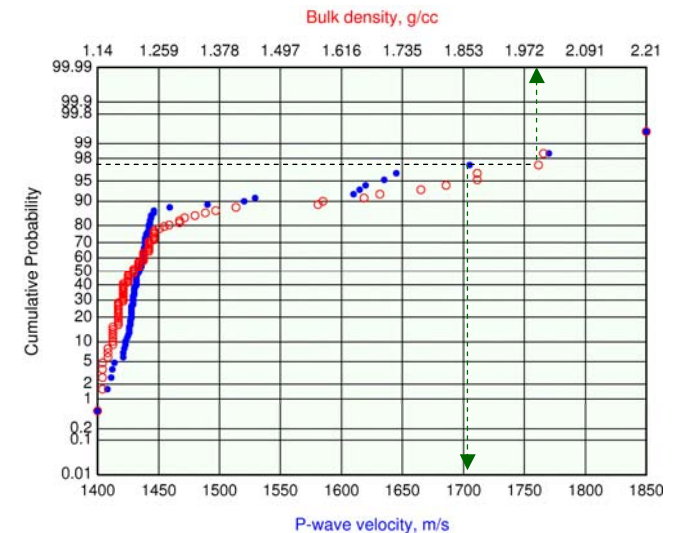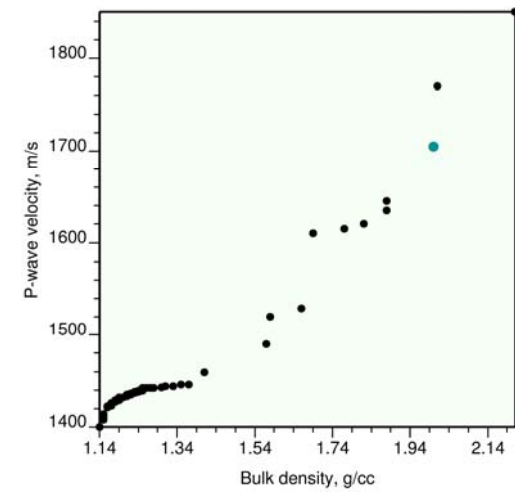
STANDARDIZED VARIATE

If X is a random variable with mean μ and standard deviation σ, the standardized variate, Z, is the transformation:

$$Z = \frac{X - \mu}{\sigma}$$

Standardized P-wave velocity

A standardized variate always has a mean of zero and variance of one.

A standardized Gaussian distribution is a called a standard normal distribution. It is often denoted by $N(0,1)$.
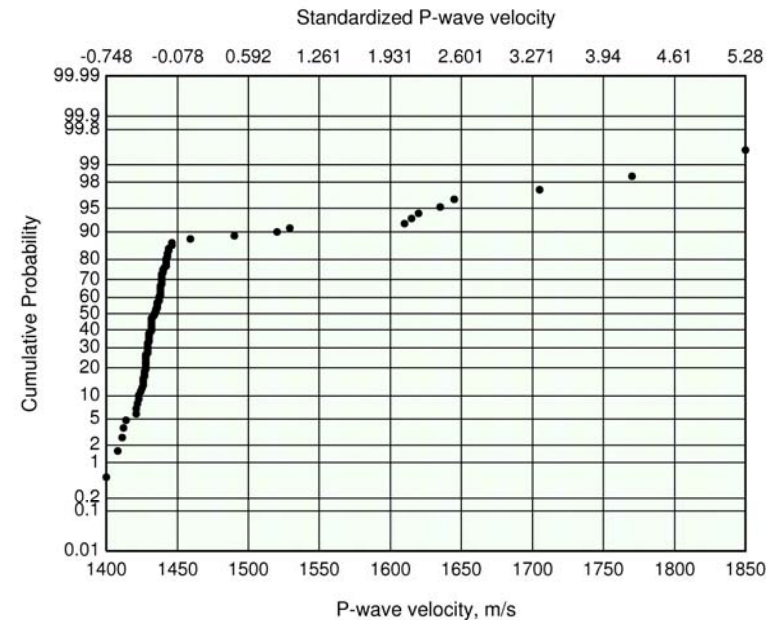
PROBABILITY-PROBABILITY (P-P) PLOT
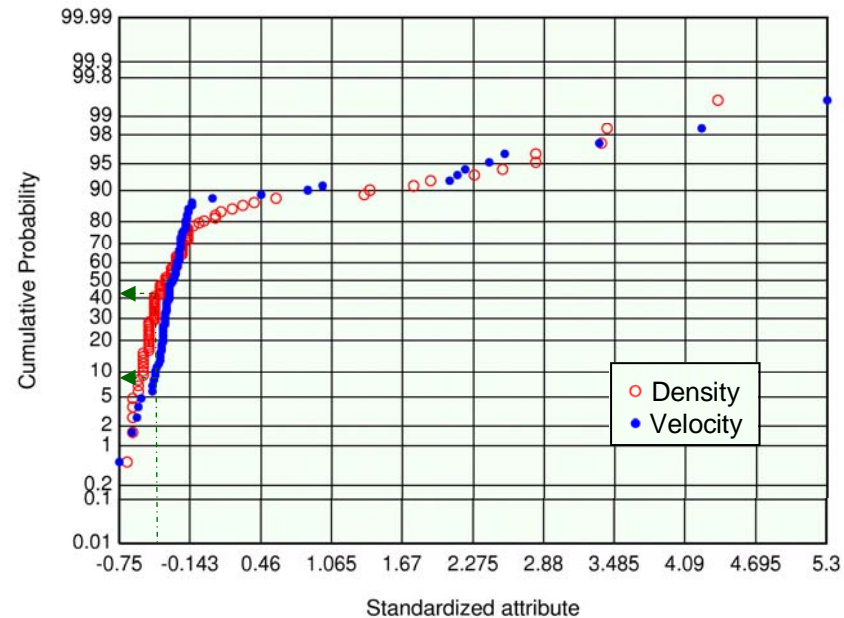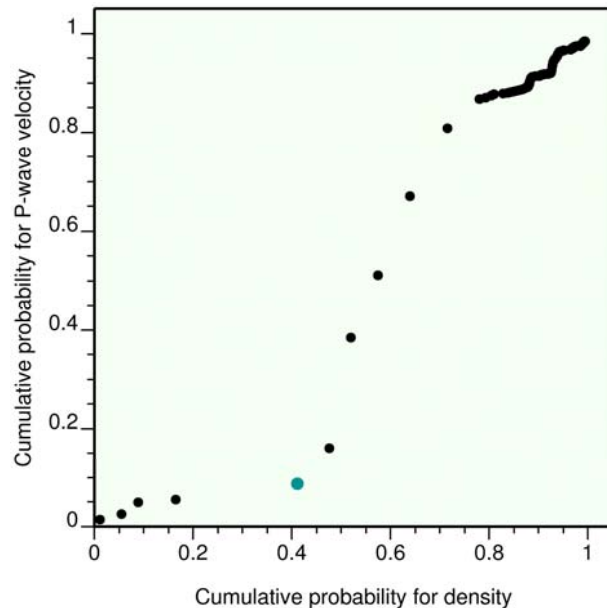
A P-P plot is another scatterplot prepared by extracting information from the cumulative distributions of two variates.

- If the variates are in different units, preliminary standardization is necessary.

- For given thresholds, the axes show the cumulative probabilities for the two distributions being compared.

Q-Q AND P-P PLOTS

Main use of the Q-Q and P-P plots is as a quick way to decide the degree of similarity between two distributions.

- If the distributions are the same, the points align along the main diagonal.

- There is no statistic or level of significance for evaluating the results.

- P-P plots are insensitive to shifting and scaling, and the vertical scale is in probability units.

- The Q-Q plot in the example illustrates its potential for calling the user's attention to the fact that normal distribution with the same parameters can take negative values.

West Lyons field, Kansas
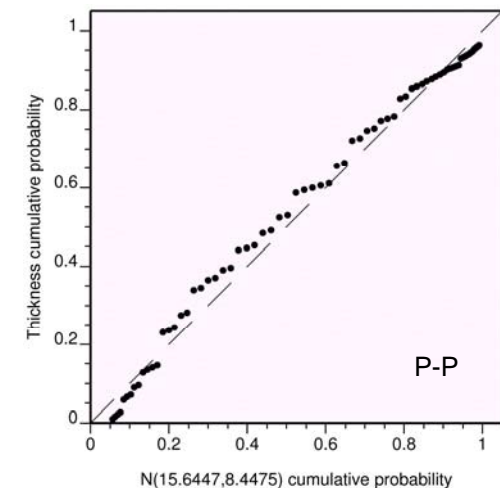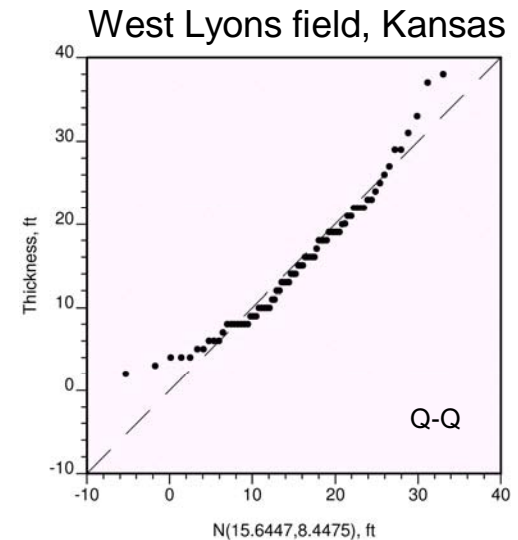
4. SGeMS SOFTWARE

GENERAL DESCRIPTION

The Stanford Geostatistical Modeling Software (SGeMS) is a general purpose, user friendly, state-of-the-art geostatistical software package.

The software code is in the public domain, downloadable from http://sgems.sourceforge.net (accessed in February 2009).

The code is in C++ and runs interactively under Windows.

A manual about to be released by Cambridge University Press is listed in the selected bibliography.

SGeMS is an outgrowth of the Stanford Geostatistical Library (GSLIB), with which it shares several features. For details about GSLIB, please consult the book listed in the bibliography.

MAIN MENU
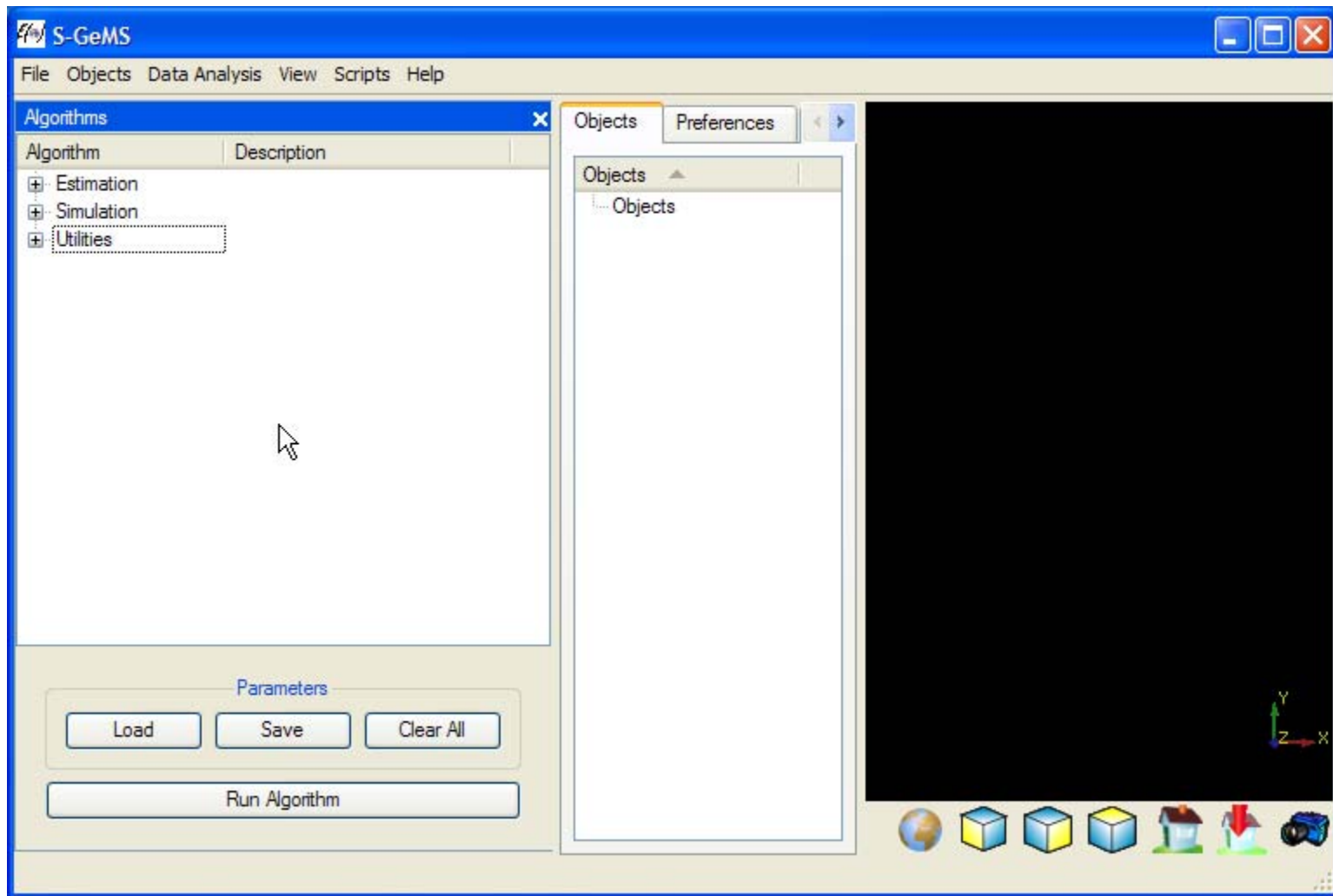
OBJECTS

Objects are files with numerical information that can be of two types:

A. Data, when each record specifies location.

B. Grid, when locations are in a regular rectangular or square pattern. Location is implicit and specified in a separate file. A grid file contains attribute values only. Typically they are the result of some calculation, but they can be used as input for subsequent processing.

Evidently, no process is possible without prior loading of an object, which is done by:

- clicking on <Objects> in the upper left corner of main menu, or

- dragging the file icon into the Object area of the main menu.

A. DATA FILES

Data may be up to three-dimensional and have to be coded into plain ASCII files. No special formatting is necessary, yet they must observe the following structure:

- The first record is intended to be a header to identify the data. It is often automatically used to label the output.

- The second record must be the number of fields per data entry.

- Starting with the third record, there should be one record per attribute containing a short description. A logical choice is to provide attribute type and unit of measurement, if any.

DATA FILES

- If the file has v attributes, the numerical values for the sample start at record $v + 3$, one for each site, under the following restrictions:

  o The entries must be strictly numeric. No letters or special symbols are allowed.

  o There must be exactly as many columns as indicated in the second record. A special code chosen by the user, such as -999, should be used if there are missing values.

  o It is required to provide at least two Cartesian coordinates specifying location.

- The number of records is unlimited.

EXAMPLE OF DATA FILE

UNCF unconformity
4
Identification
Easting, m
Northing, m
Depth, ft

1	32000	87015	7888
2	38600	87030	8020
3	44000	86400	7949
⋮			
6200	85050	87000	8003
6202	49400	84000	8032

DATA LOADING

Simply:

- Drag the file icon into the object area of the main menu

or:

- Click on <Objects>.

- Click on <Load Object> at new menu.

- Search for the file of interest.

- Highlight file name and click on <OK>.

The top of the file will be displayed. Finish by:

- Choosing <Point Type> at <Select object type>.
- Clicking on <Next>.

1. COMPLETING DATA LOADING

- Type an object name at <Pointset name>.
- Indicate columns for easting and northing and also depth in three-dimensional studies.
- Specify missing data code, if any.
- Click on <Finish>.

The loading has been completed and the data file will show in the object panel of the main menu.

B. GRID LAYOUT
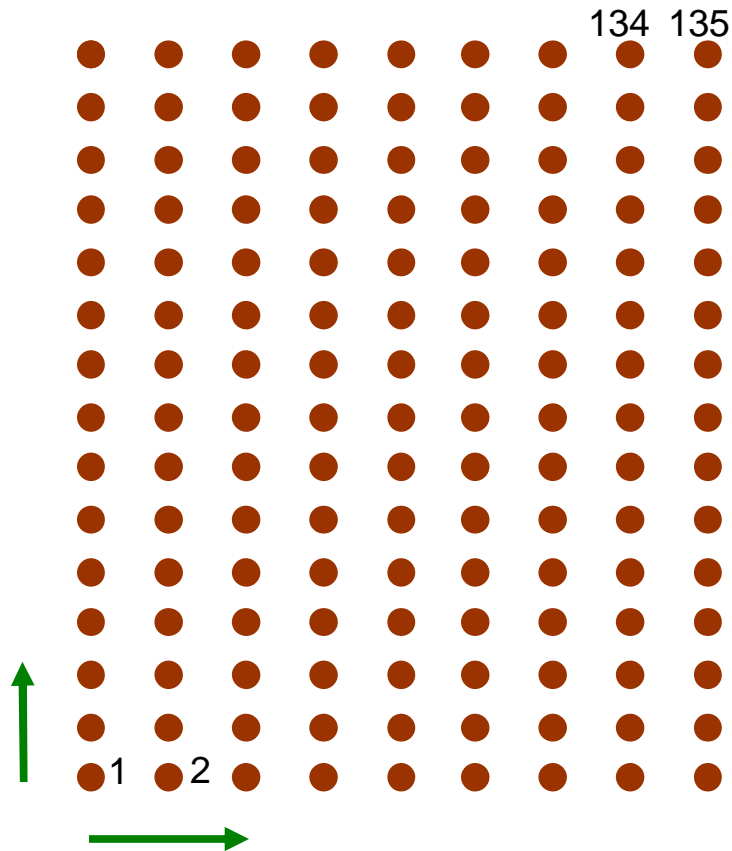
Results from modeling of spatial attributes come in the form of arrays of values at regular intervals.

Coordinates for the nodes are not printed; they are implicit in the regular arrangement.

IMPLICIT ARRANGEMENT

- In three-dimensional modeling, layer values go from the lowermost to uppermost layer.
- Within a layer, the first value to go into the file is the one in the lower left corner.
- Then come the rest of the values for the lowermost row, from left to right.
- Within a layer, last value to enter the file is the one in the extreme upper right corner.
- Specification of grid geometry goes into a separate file.

GRID SPECIFICATION FILE

Remember that grids must be specified before attempting any processing that will result in the generation of a grid, such as estimation or simulation.

EXAMPLE OF GRID

To minimize storage space, SGeMS stores grid values in binary form. Grids can only be displayed in graphical form.

Numerical display of a grid requires saving it in ASCII format and further viewing it with a text editor. To do such saving:

- Click on <Object>.
- Select <Save Object>.
- Name the file.
- Select place for storage at <Look in>.
- Make sure to select GSLIB as the format.
- Click on <Save>.
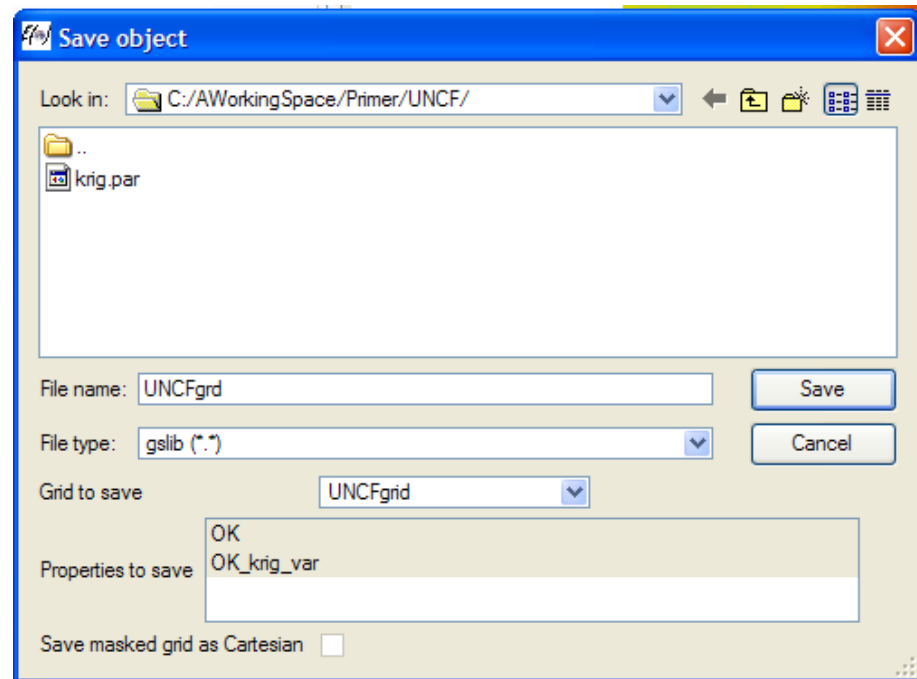
VIEW OF GRID WITH TEXT EDITOR

UNCFgrid (211x281x1)
2
OK
OK_krig_var
7773.04   2075.6
7766.52   1902.38
7763.12   1745.86
7759.76   1595.19
7756.46   1450.97
         ⋮
8017.93   2229.09
8018.47   2313.04
8018.98   2404.05
8019.46   2501.91
8019.92   2606.35

SGeMS automatically creates the first record, which includes the number of columns, rows, and layers.

OBJECT SAVE

In SGeMS, all objects in a session make up a project.

The user can save all objects in a project for archiving or further use.
To do that:

- Click on <File> at the upper right corner of the SGeMS screen;
- Click on <Save Project>. This is all that it is necessary to do for an existing project.
- If the project is new, navigate directories to locate a storage folder, select name for the project file, and click on <Save>.
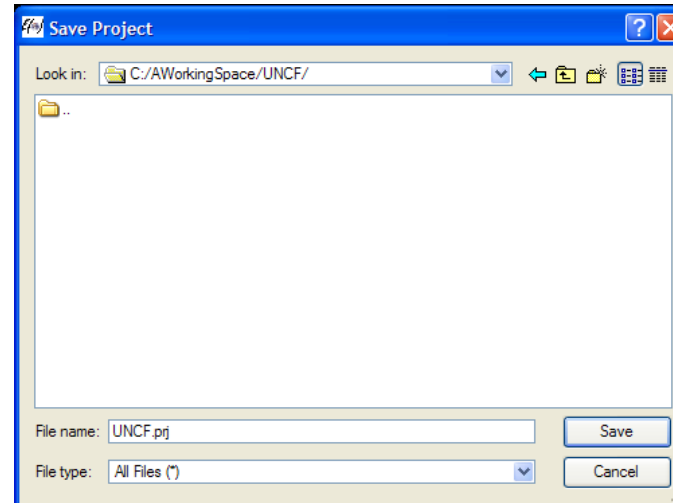
HISTOGRAM PREPARATION

As a first practical application of SGeMS, let us prepare a histogram of the object previously loaded. For that:

- Click on <Data Analysis> at main menu.

- In the new menu, click on <Histogram>.

- Select the object.

- Choose the property.

- Set the number of classes.

- If there is need to change the axis extreme values, click on <Display Option> and do the settings.
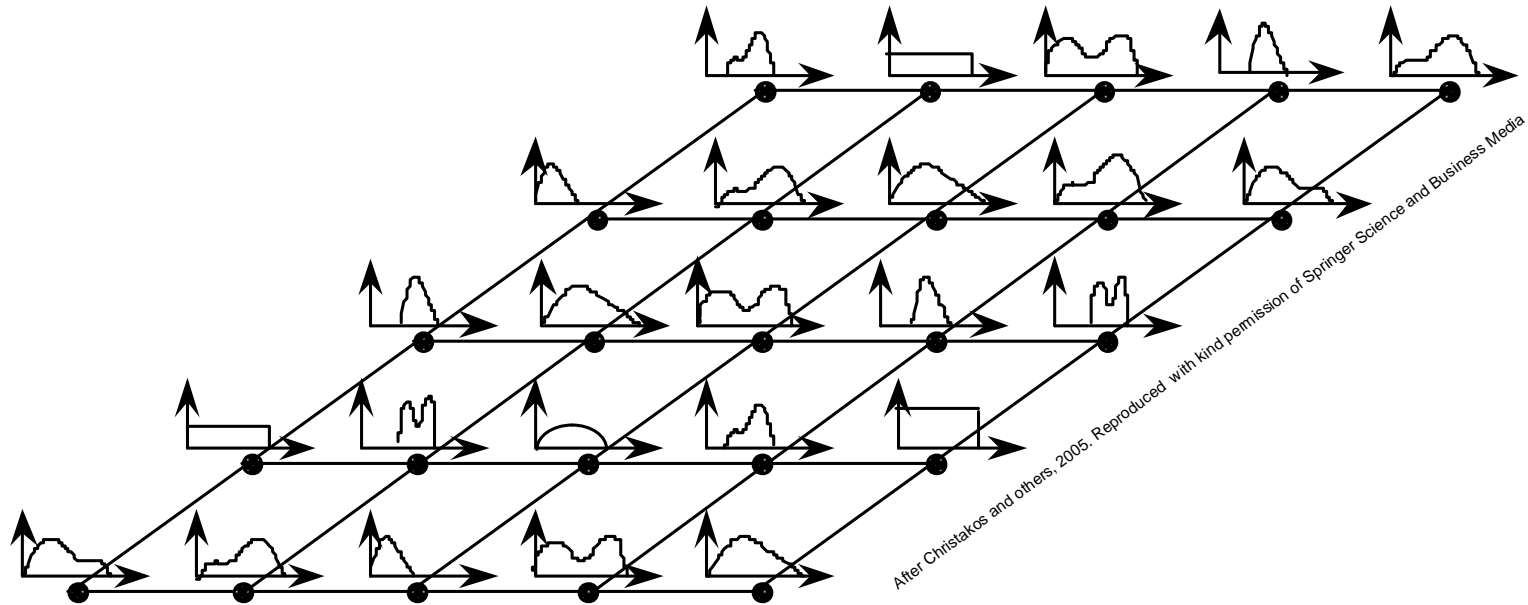
HISTOGRAM DISPLAY
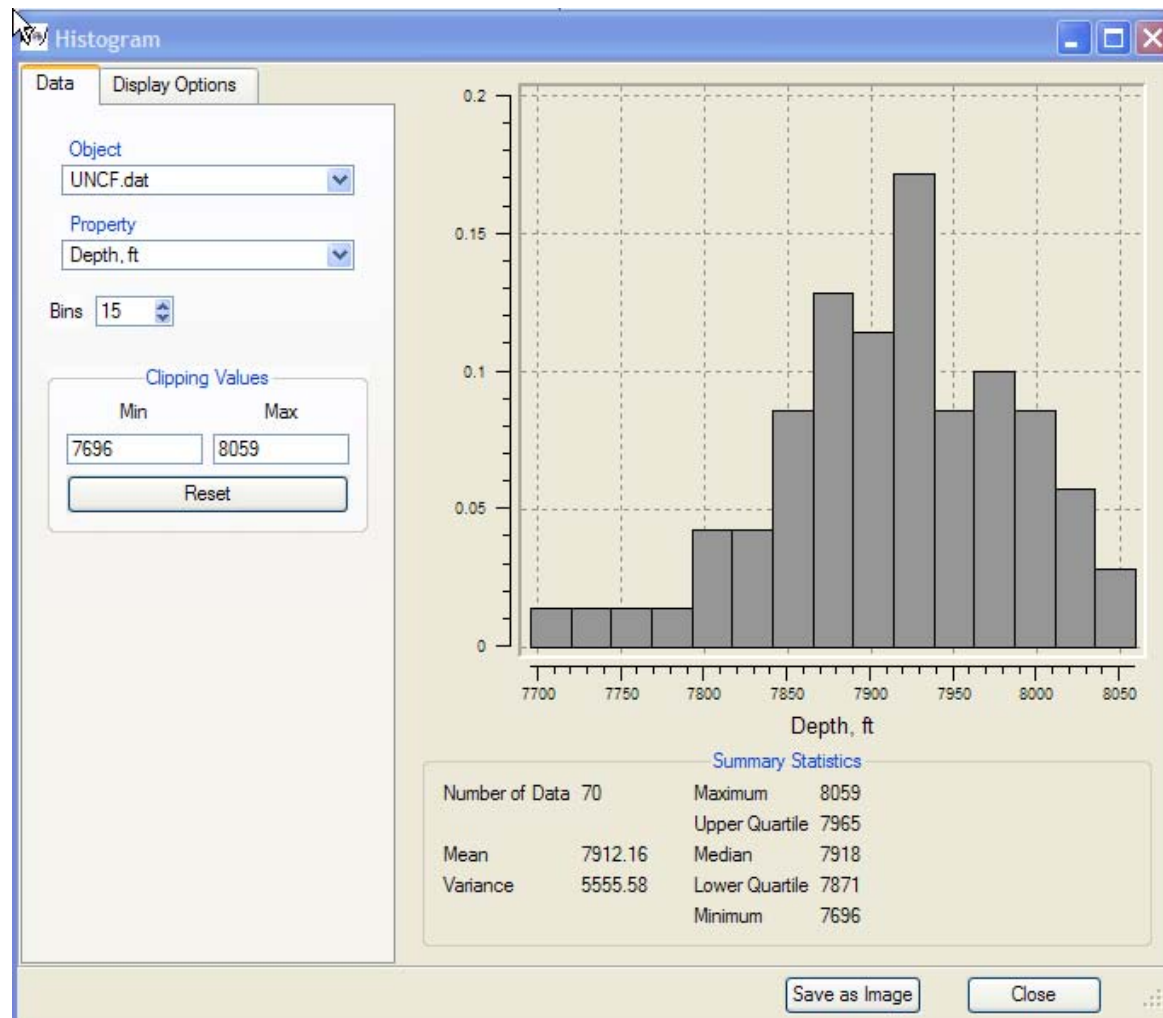
SAVING IMAGE

- Click on <Save as Image>.
- Select a folder.
- Give a name to the file.
- Select type.
- Decide on grid and summary statistics.
- Click on <Save>.
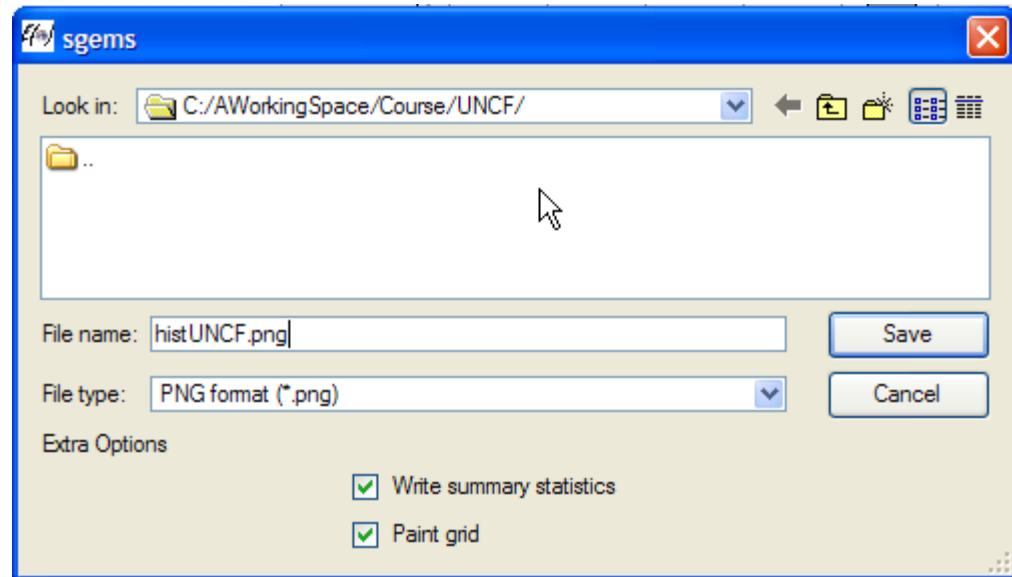
The histogram graphical file is now at the specified folder.

5. SPATIAL STATISTICS

RANDOM FUNCTIONS

After Christakos and others, 2005. Reproduced with kind permission of Springer Science and Business Media

A random function is a collection of random variables, one per site of interest in the sampling space.

A realization is the set of values that arises after obtaining one outcome for every distribution.

Geostatistics relies heavily on random functions to model uncertainty.

MAIN DIFFERENCES BETWEEN CLASSICAL GEOSTATISTIC AND CLASSICAL STATISTICS

- Geographical location is an integral part of any sample.
- Measurements have an associated support. For example, in the case of density, the support is the specimen volume.
- Geostatistics does not assume that the variables are independent and identically distributed.
- A spatial sample is regarded as a single realization of a random function, instead of multiple outcomes of a single random variable.
- Most geostatistical formulations do not require any particular probability distribution, although some work better under normality conditions.
- The most adequate forms of spatial sampling are those following a regular pattern—cubic or hexagonal, for example.

ERGODIC ASSUMPTION

We have seen that stochastic methods regard data as one of an infinite number of possible realizations.

Ordinarily, one needs multiple realizations to infer properties of the complete ensemble.

The ergodic assumption basically states that just one realization is indeed sufficient to make reliable assessments about ensemble properties.

It is an assumption because it is not possible to test it.

STATIONARITY

This is another fundamental assumption, presuming invariance through space of properties of the random function.

The most general case assumes that the joint probability distribution of any number of random variables is the same.

Most geostatistical formulations require independence of location only for moments and just up to order two.

BIAS

Bias in sampling denotes preference in taking the measurements.

Bias in an estimator implies that the calculations are preferentially loaded relative to the true value, either systematically too high or too low.

• When undetected or not compensated, bias induces erroneous results.
• The opposite of the quality of being biased is to be unbiased.

PREFERENTIAL SAMPLING

Although estimation and simulation methods are robust to clustered preferential sampling, parameters that need to be inferred from the same sample prior to estimation or simulation can be seriously distorted, especially for small samples.

The solution to preferential sampling is preparation of a compensated sample to eliminate the clustering, for which there are several methods.

Declustering is important for the inference of global parameters, such as any of those associated with the histogram.

EXAMPLE OF SAMPLE WITH PREFERENTIAL CLUSTERING

Clustered sample
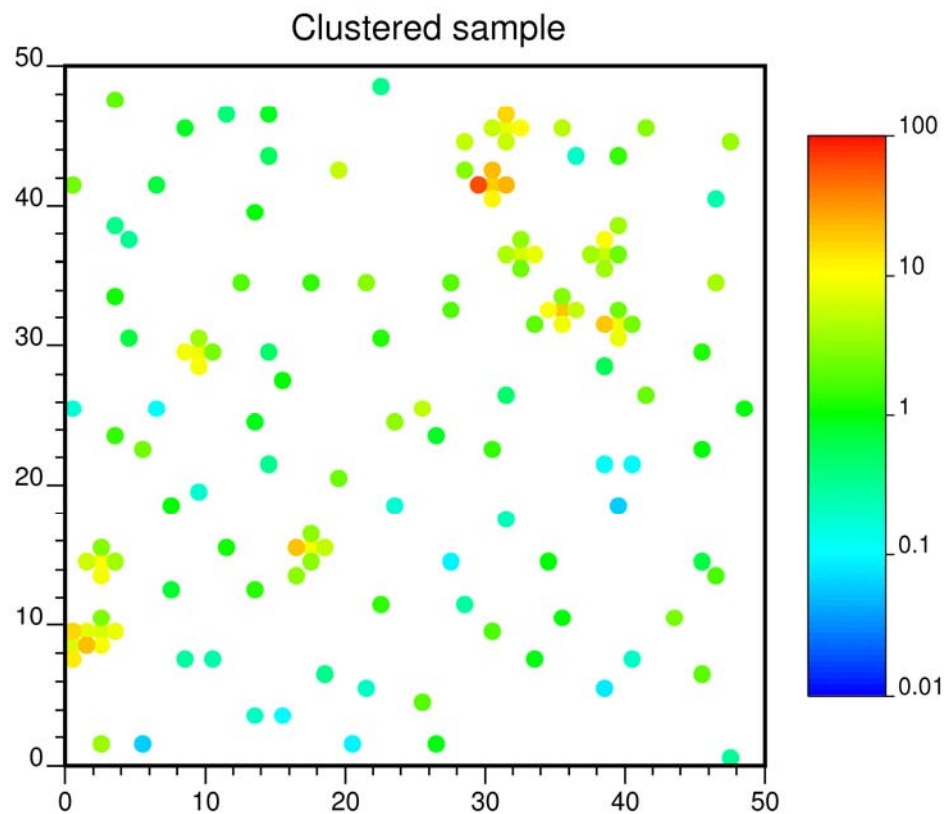
DECLUSTERING (1)

Detect presence of clusters by preparing a cumulative distribution of distance to nearest neighbor.

DECLUSTERING (2)
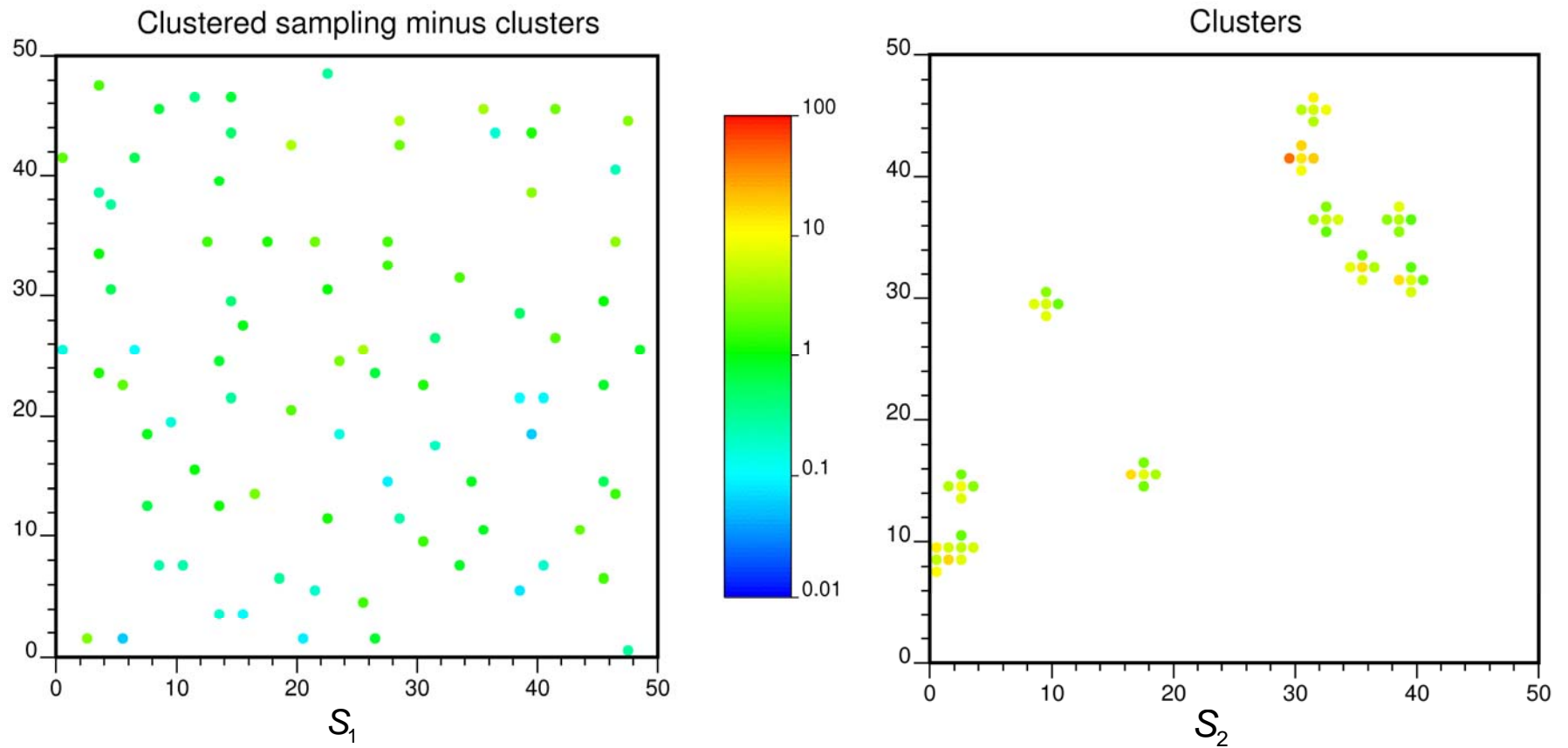
Decompose the clustering.

DECLUSTERING (3)

Histograms of the attribute by distance class
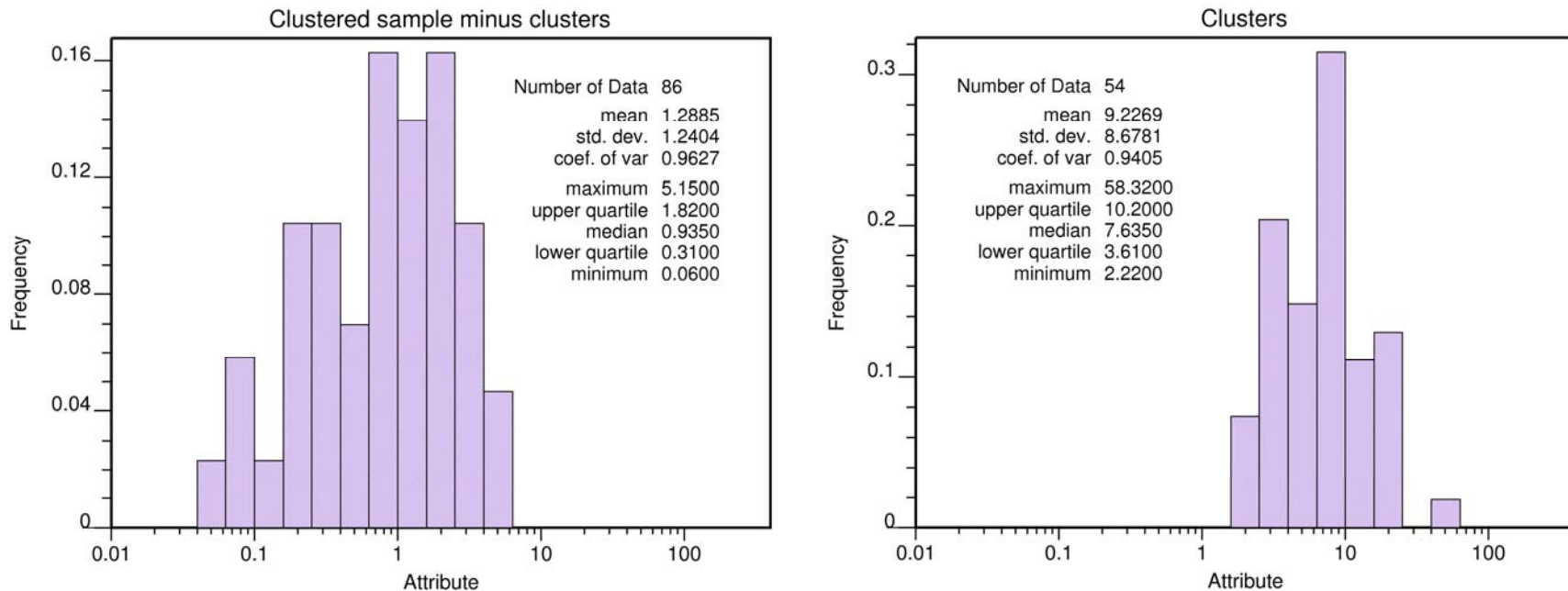
Preferential sampling shows as poor overlapping between the two histograms.

DECLUSTERING (4)

One possibility is to obtain the declustered subset ($S4$) by expanding the subset without clusters ($S1$) by transferring a few observations (S_3) from the clustered subset (S_2).

Distance to nearest neighbor, cluster-free subset S_1 (y-axis)

Distance to nearest neighbor, transfer subset S_3 (x-axis)
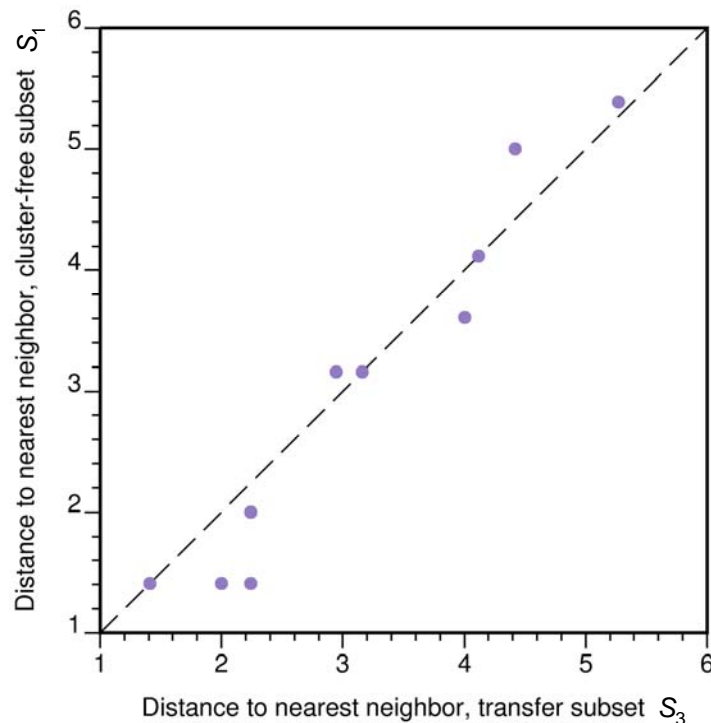
- Transfer observations from (S_2) by decreasing distance to nearest neighbor in S_4.
- Stop transferring points when the distribution of distances for the original subset S_1 is about the same as that for the transferred observations (S_3) within S_4.

POSTING OF THE DECLUSTERED SAMPLE (S_4)
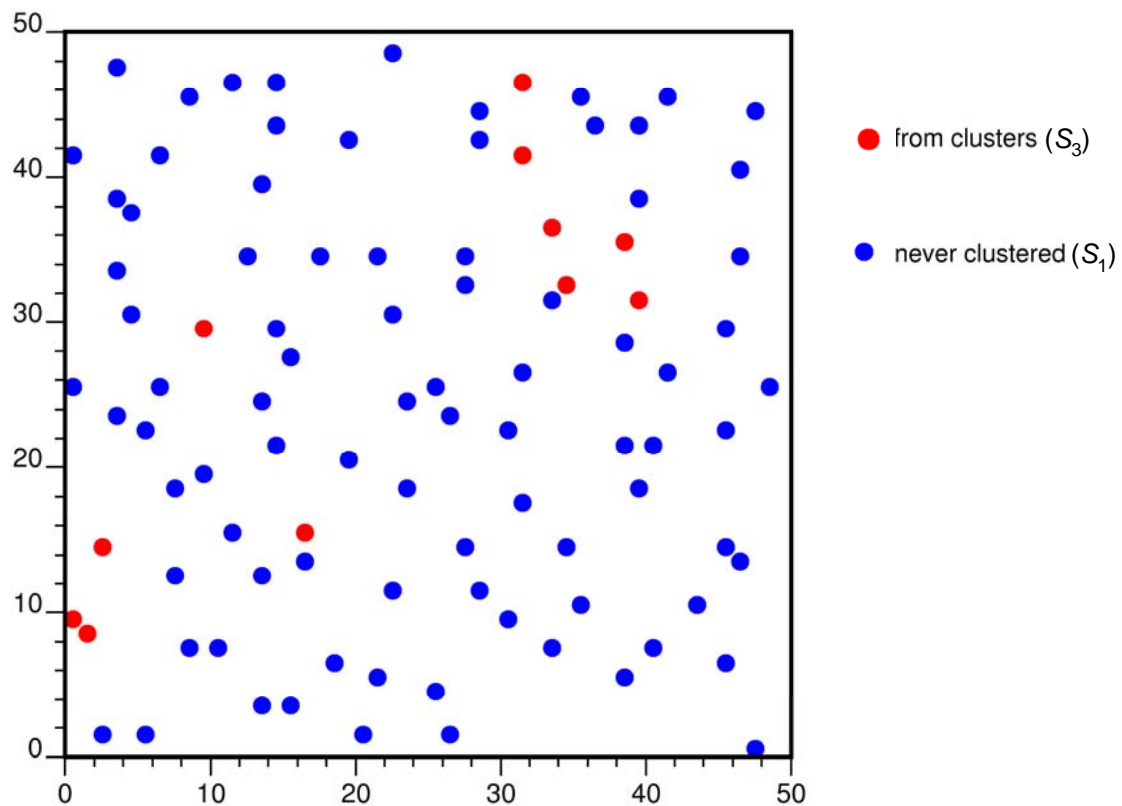
6. TRANSFORMATIONS

COMPARING DISTRIBUTIONS

The distribution followed by some data and the one that may be required by a method may be two different distributions.

If a method requires a particular distribution, such distribution becomes mandatory.

The simplest way to compare a sample, $F_n(x)$, to an hypothesized distribution, $F(x)$, is to superimpose their cumulative distributions.

By transforming the data, one can force the data to follow almost any other distribution.
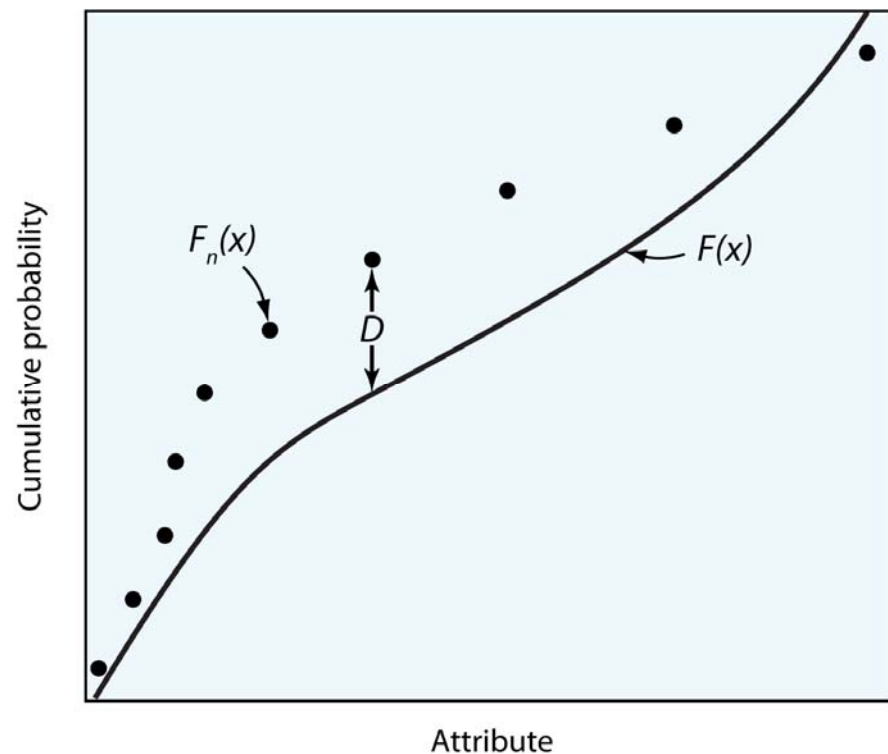
KOLMOGOROV-SMIRNOV TEST

The maximum discrepancy, *D*, is the measure of similarity in the Kolmogorov-Smirnov test.

The significance of the discrepancy depends on the sample size and on whether the parameters of the hypothesized distribution are known. A statistic called *p*-value, commonly read from tables, is used to decide the significance of the discrepancy. Customarily, if the *p*-value is above 0.1 or 0.05, the equality of the distributions is accepted.

Standard tables available to run the significance of the discrepancy are not valid in geostatistics because the tables do not take into account spatial dependence.

It has been found, however, that the standard tables may be a limiting case for the situation considering spatial correlation. In that case, each time two distributions pass the Kolmogorov-Smirnov test without considering spatial correlation, they also pass it considering that correlation. The same is not true in case of rejection.

WEST LYONS FIELD, KANSAS

Remember that the cumulative probability of any normal distribution follows a straight line in normal probability scale.

At 10 ft, the thickness distribution of Lyons field reaches a maximum discrepancy of 0.09 against a normal distribution with the same mean and standard deviation as the sample. The p-value without considering spatial correlation, p_u, is 0.14.

Because the p_u-value is already above the threshold of 0.1–0.05, it is not necessary to go into the complications of considering spatial correlation. Without further ado, it can be accepted that thickness is normally distributed.
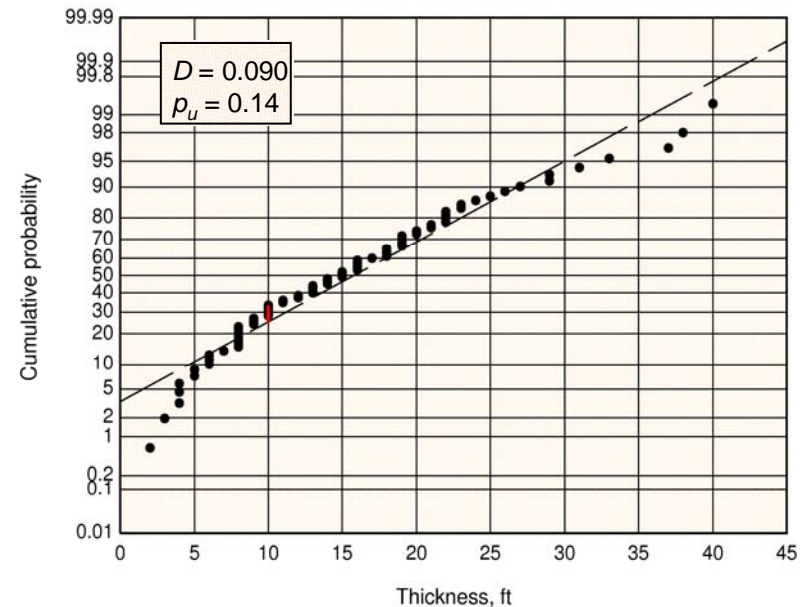
NORMAL SCORES

In geostatistics, as in classical statistics, the most versatile distribution is the normal distribution.

If the sample strongly deviates from normality, it may be convenient or required to transform the data to follow a normal distribution, so as to properly apply a specific method.

The transformation is fairly straightforward. It is performed by assigning to the cumulative probability of every observation the value of the standard normal distribution for the same cumulative probability.

NORMAL SCORES
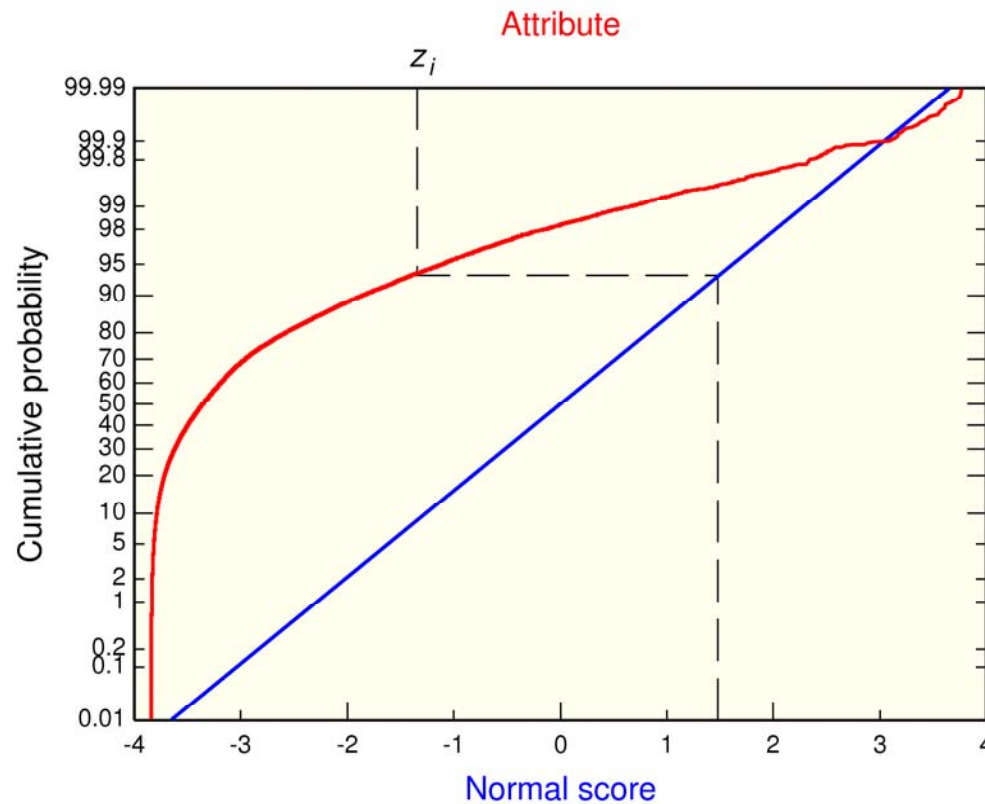
For the value z_i of the attribute, the normal score is 1.45.

   The transformation also can be used backwards to bring calculated values in the normal score space to the original attribute space.

LOGNORMAL TRANSFORMATION

- This is a special case of normal score transformation. The transformed value is simply the logarithm of the original observation.

- If the original distribution is exactly lognormal, the distribution of the transformed values will be exactly normal. Because this is never the case, the distribution for a transformed variable is only approximately normal.

- Most minerals and chemical elements have distributions of concentration close to lognormal, which explains the great attention paid to this distribution since the early days of geostatistics.

INDICATORS

Given a continuous random function $Z(\mathbf{s})$, its indicator $I(\mathbf{s}, u)$ is the binary transformation:

$$I(\mathbf{s}, u) = \begin{cases} 0, \text{ if } Z(\mathbf{s}) > u \\ 1, \text{ if } Z(\mathbf{s}) \leq u \end{cases}$$

The potential of indicators lies in the possibility to handle imprecise data and the fact that the expected value of an indicator is equal to the cumulative probability for the threshold value.

$$E[I(\mathbf{s}, u \mid (n))] = \text{Prob}[Z(\mathbf{s}) \leq u \mid (n)]$$
$$= F(\mathbf{s}, u \mid (n))$$

INDICATOR CODING FOR DIFFERENT TYPES OF CONTINUOUS DATA
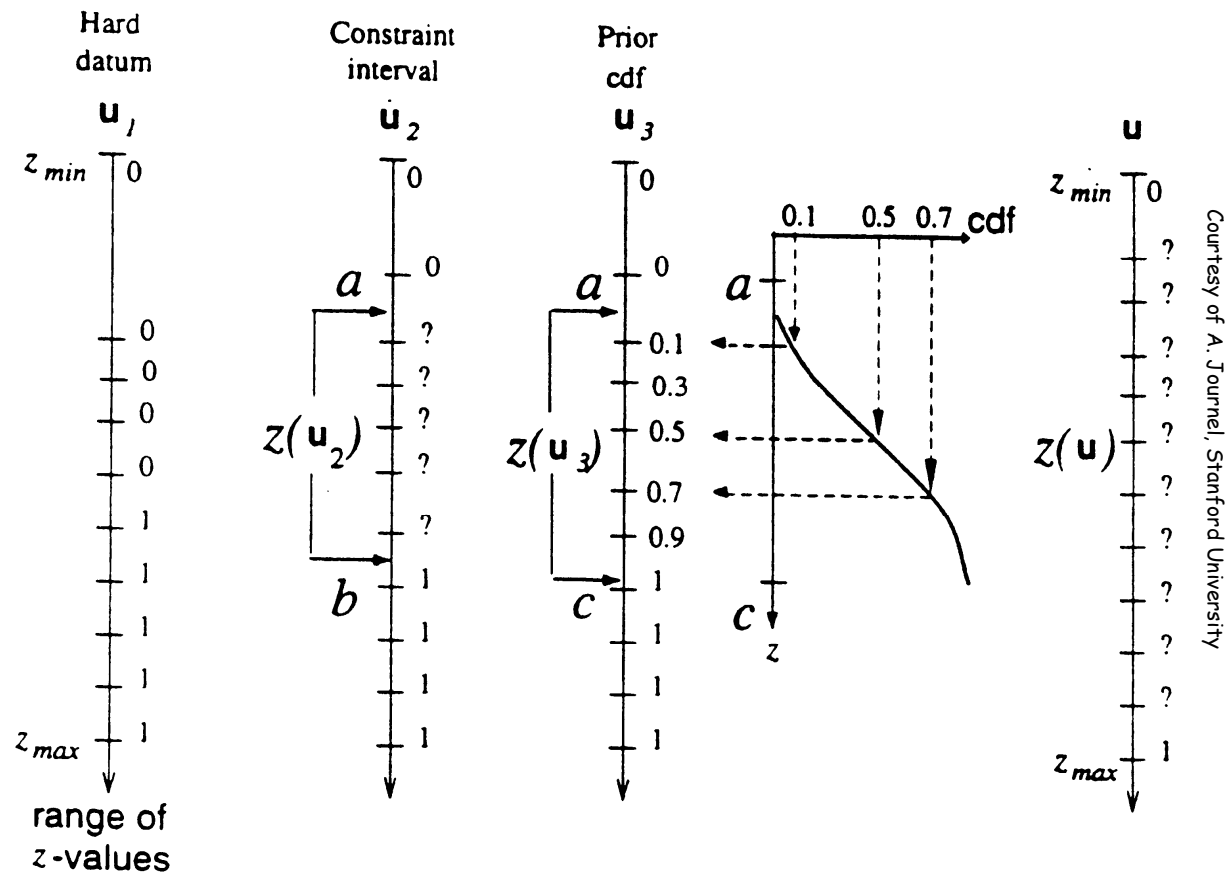
DISCRETE INDICATORS

The indicator $I(\mathbf{u}, k)$ of a categorical attribute is:

$$I(\mathbf{u}, k) = \begin{cases} 1, \text{if } u \in k \\ 0, \text{otherwise} \end{cases}$$

In this case, the indicator is now equal to the probability of belonging to class k.

$$E[I(\mathbf{u}, k \mid (n))] = \text{Prob}[u \in k \mid (n)]$$

In this case the indicator transformation offers the possibility of digital processing for information that is commonly qualitative.

7. SEMIVARIOGRAM

THE OBJECTIVE
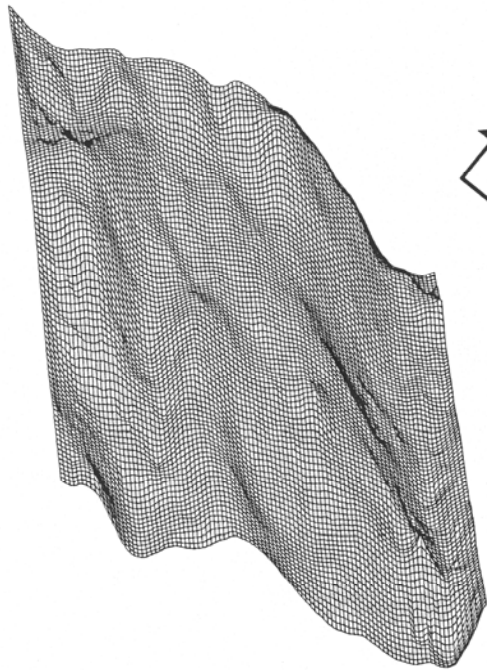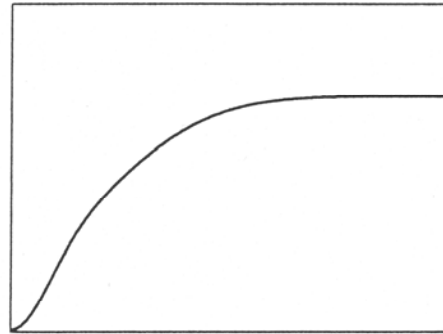
The objective of structural analysis is to capture the style of the fluctuations through an analytical function.

109

CONTINUITY

Spatial continuity involves the concept that small values of an attribute are in geographical proximity to other small values, while high values are close to other high values. Transitions are gradual.

Assessment of covariance or its close equivalent, the semivariogram, has been the classical way in geostatistics to measure spatial correlation.

RELEVANCE

- The semivariogram or the covariance are necessary in the formulation and application of most estimation and simulation methods.

- Semivariogram analysis leads to some stand-alone applications, such as sampling design.

USGS

DEFINITIONS

Given a spatial distance **h** involving magnitude and direction, the semivariogram $\gamma(\mathbf{h})$ is:

$$\gamma(\mathbf{h}) = \frac{1}{2} Var\left[Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h})\right]$$

and the spatial covariance is:

$$\mathrm{Cov}(\mathbf{h}) = \mathrm{E}\left[(Z(\mathbf{s}) - m)(Z(\mathbf{s} + \mathbf{h}) - m)\right],$$

both with a constant mean m

$$m = \mathrm{E}\left[Z(\mathbf{s})\right] = \mathrm{E}\left[Z(\mathbf{s} + \mathbf{h})\right].$$

   Differently from the classical covariance (page 56), the spatial covariance is grouped by distance and applies to the same attribute. Therefore a more descriptive yet seldom used term for the covariance is autocovariance.

EQUIVALENCE

If the mean is constant and the covariance is independent of location, then always

$$\gamma(\mathbf{h}) = \mathrm{Cov}(0) - \mathrm{Cov}(\mathbf{h}),$$

which makes it immaterial which one to use.

In general, what is estimated is the semivariogram because its estimation does not require knowledge of the mean.

Lag *a* is the range and the semivariogram asymptote is called the sill, which is equal to the variance Cov(0).
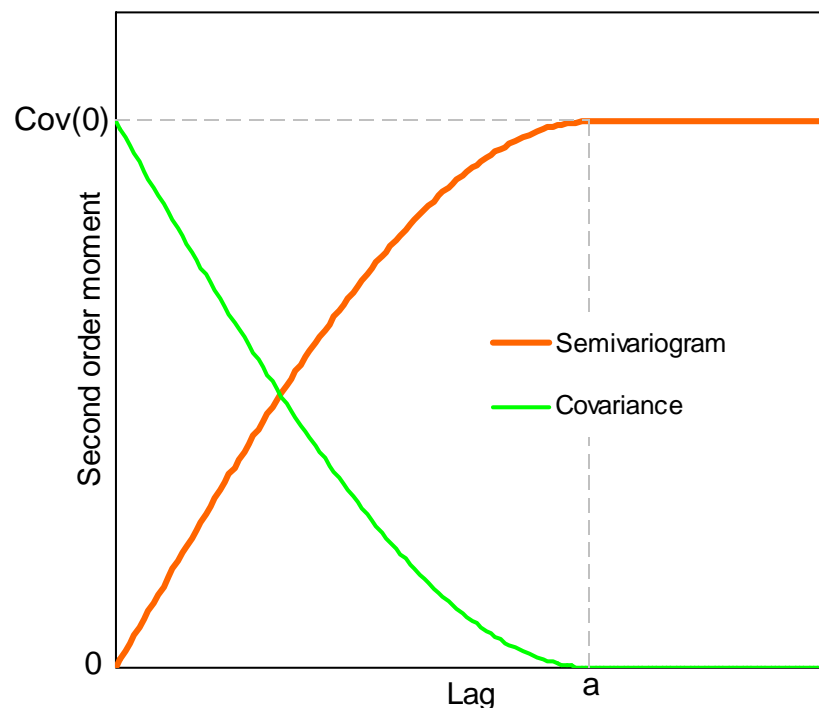
SEMIVARIOGRAM ESTIMATOR

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2n(\mathbf{h})} \sum_{i=1}^{n(\mathbf{h})} [z(\mathbf{s}_i) - z(\mathbf{s}_i + \mathbf{h})]^2$$

$\hat{\gamma}(\mathbf{h})$: estimated or experimental semivariogram;

$\mathbf{h}$: lag;

$n(\mathbf{h})$: number of pairs $\mathbf{h}$ units apart;

$z(\mathbf{s}_i)$: observation at site $\mathbf{s}_i$.

Always remember that for the estimator to be valid, the mean must be constant; hence, there must be no trend.

For any direction, it is recommended that $n(\mathbf{h}) \geq 30$ and the estimation be restricted to $\mathbf{h}$ no more than half the extension of the sampling domain in such direction.

TRIVIAL EXAMPLE

$\mathbf{s}_i$	$z(\mathbf{s}_i)$	Δ^2_0	$\Delta^2_{0.5}$	$\Delta^2_{1.0}$	$\Delta^2_{1.5}$	$\Delta^2_{2.0}$	$\Delta^2_{2.5}$	$\Delta^2_{3.0}$	$\Delta^2_{3.5}$	$\Delta^2_{4.0}$	$\Delta^2_{4.5}$
7.0	3.2	0.0	1.21	3.24	10.89	22.09	24.01	18.49	16.81	12.25	3.76
7.5	4.3	0.0	0.49	4.84	12.96	14.44	10.24	9.00	5.76	2.25	
8.0	5.0	0.0	2.25	8.41	9.61	6.25	5.29	2.89	0.64		
8.5	6.5	0.0	1.96	2.36	1.0	0.64	0.04	0.49			
9.0	7.9	0.0	0.04	0.16	0.36	1.44	4.41				
9.5	8.1	0.0	0.36	0.64	1.96	5.29					
10.0	7.5	0.0	0.04	0.64	2.89						
10.5	7.3	0.0	0.36	2.25							
11.0	6.7	0.0	0.81								
11.5	5.8	0.0									
$\sum \Delta^2_{\mathbf{h}}$		0.0	7.52	22.74	38.97	50.15	43.99	30.87	23.21	14.50	6.76
$n(\mathbf{h})$		10	9	8	7	6	5	4	3	2	1
$\lambda(\mathbf{h})$		0.0	0.42	1.42	2.78	4.18	4.40	3.85	3.86	3.62	3.38
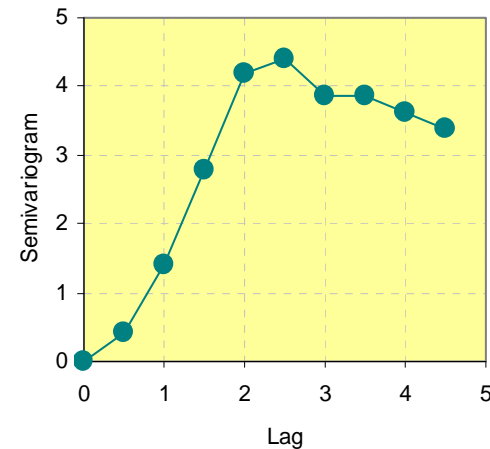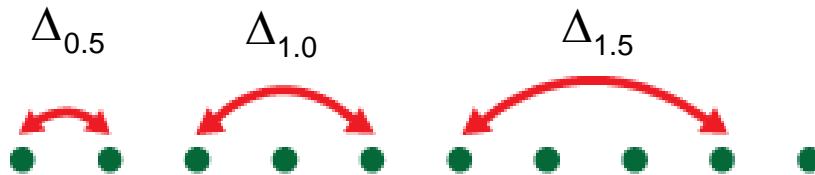
CASE OF IRREGULAR PATTERNS

If the observations, $z(\mathbf{s})$, are not regularly spaced, they are grouped into distance classes of equal radial thickness, which customarily is set equal to twice t_h.
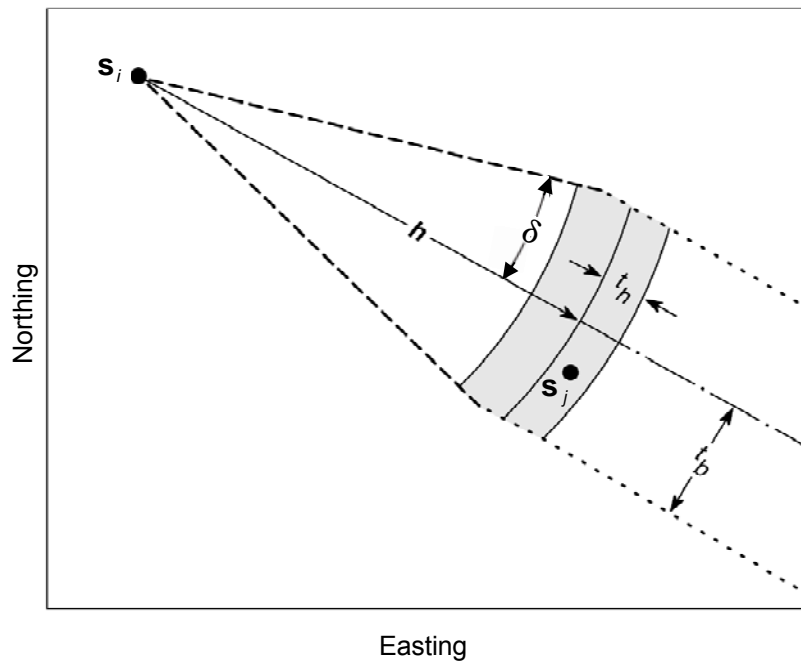
Commonly used parameters are:

t_h : lag tolerance

δ : angular tolerance

t_b : bandwidth

t_b only operates when the cone is wider than twice t_b. Otherwise, δ prevails.

Each class contributes one value to the experimental semivariogram, $\hat{\gamma}(\mathbf{h})$.

MOST COMMON PERMISSIBLE MODELS

Permissible semivariogram models are a class of models guaranteeing unique solution to an estimation system of equations and a nonnegative estimation variance.

Spherical

$$\gamma(\mathbf{h}) = C\left[\frac{3}{2}\frac{\mathbf{h}}{\mathbf{a}} - \frac{1}{2}\left(\frac{\mathbf{h}}{\mathbf{a}}\right)^3\right]$$

Exponential

$$\gamma(\mathbf{h}) = C\left[1 - e^{-3\left(\frac{\mathbf{h}}{\mathbf{a}}\right)}\right]$$

Gaussian

$$\gamma(\mathbf{h}) = C\left[1 - e^{-3\left(\frac{\mathbf{h}}{\mathbf{a}}\right)^2}\right]$$
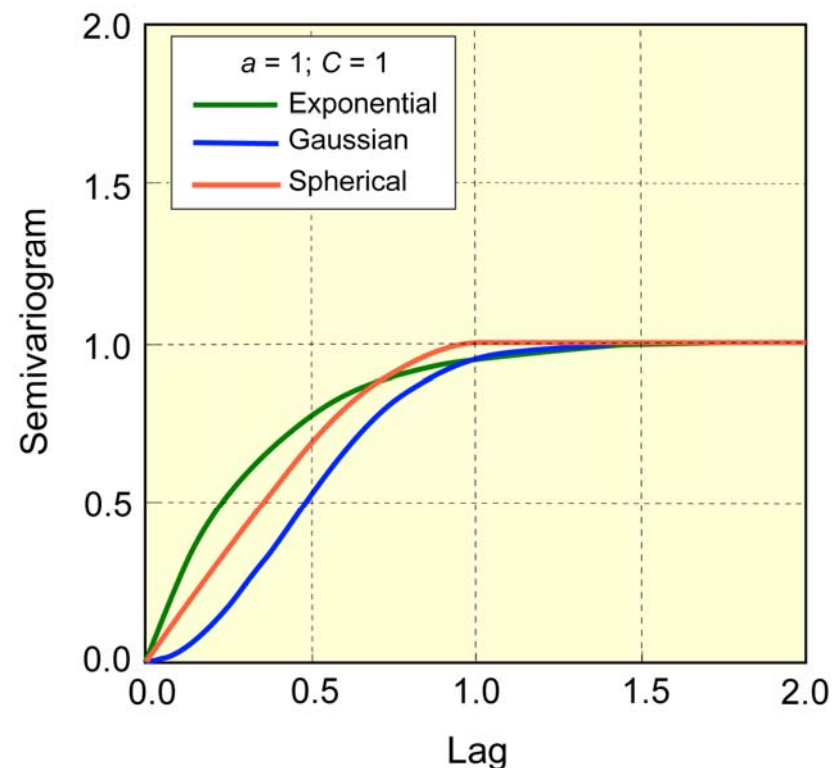
A sum of permissible models is also permissible.

ASSESSMENT OF SPATIAL CORRELATION

Spatially correlated attributes have gradually increasing semivariograms that reach the range at a distance clearly greater than zero.

If the semivariogram does not seem to be zero for a lag close to zero, the value of convergence is called nugget effect. Yet, $\gamma(0) = 0$.

Lack of spatial correlation is revealed by a semivariogram that remains constant, following what is called a pure nugget-effect model, which is permissible.

Mere assessment of spatial correlation does not require semivariogram modeling.

Proof of spatial correlation

Nugget effect

Lack of spatial correlation
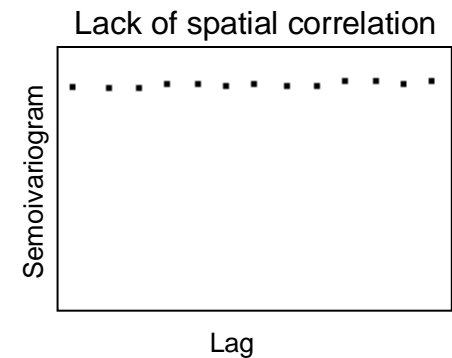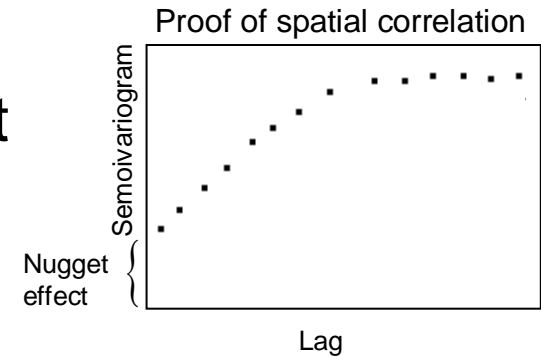
ESTIMATION BASIC STEPS

A. Make sure that the data are error free.

B. Correct preferential clustering, if any.

C. Perform transformations if required or desired.

D. Estimate semivariograms along at least 3 different directions.

E. Detect trend. In the presence of trend, the experimental semivariogram is an artifact that increases monotonically, reaching values orders of magnitude higher than the sample variance.

F. If necessary, model semivariogram along the trend-free direction.

   When the attribute is isotropic, parameters do not change with **h**. A semivariogram model prepared disregarding azimuthal variation is said to be omnidirectional.

A. DATA LOADING
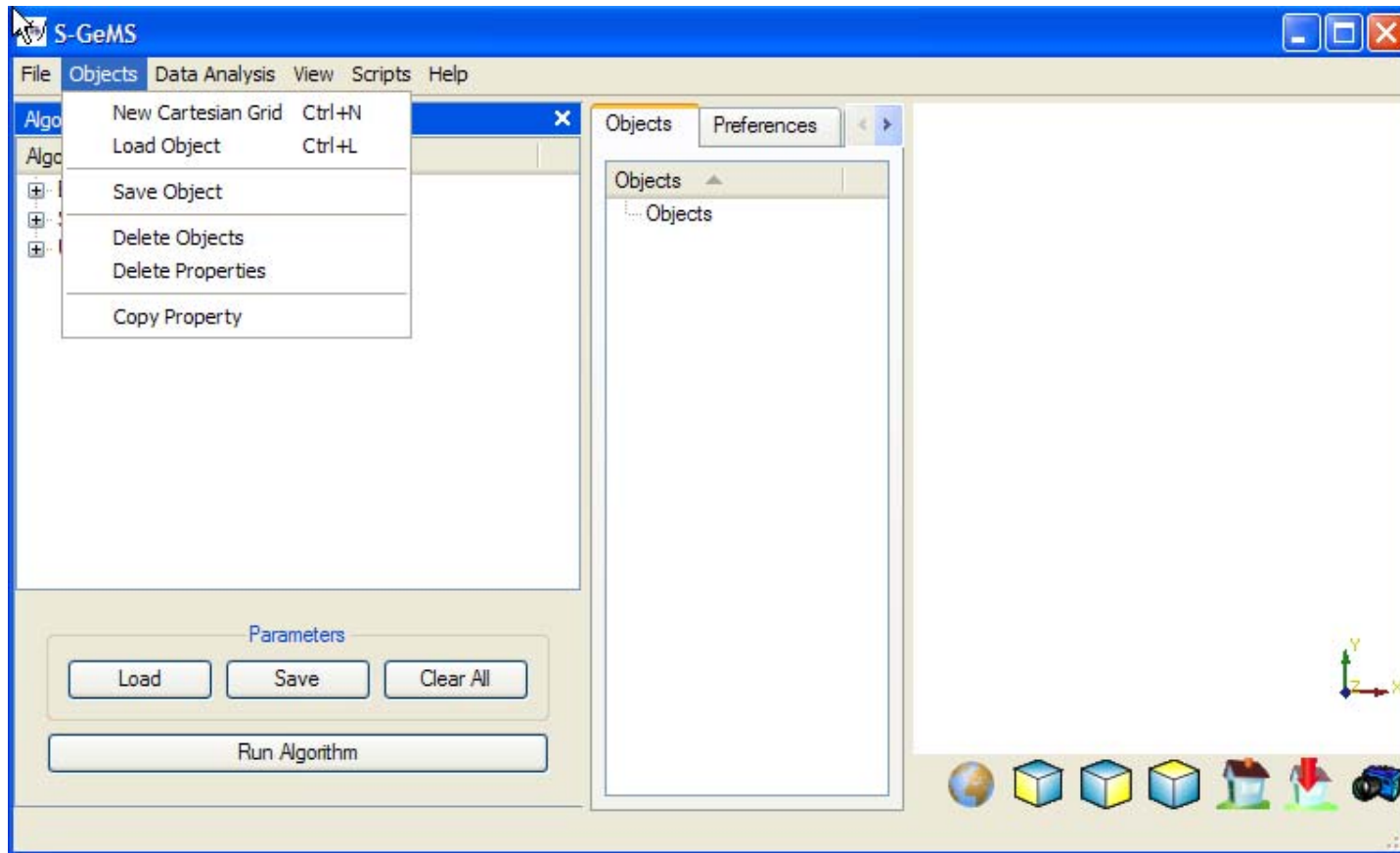
If data have not yet been loaded into SGeMS, load them as explained in chapter 4 (page 74).

B. MODELING WITH SGeMS
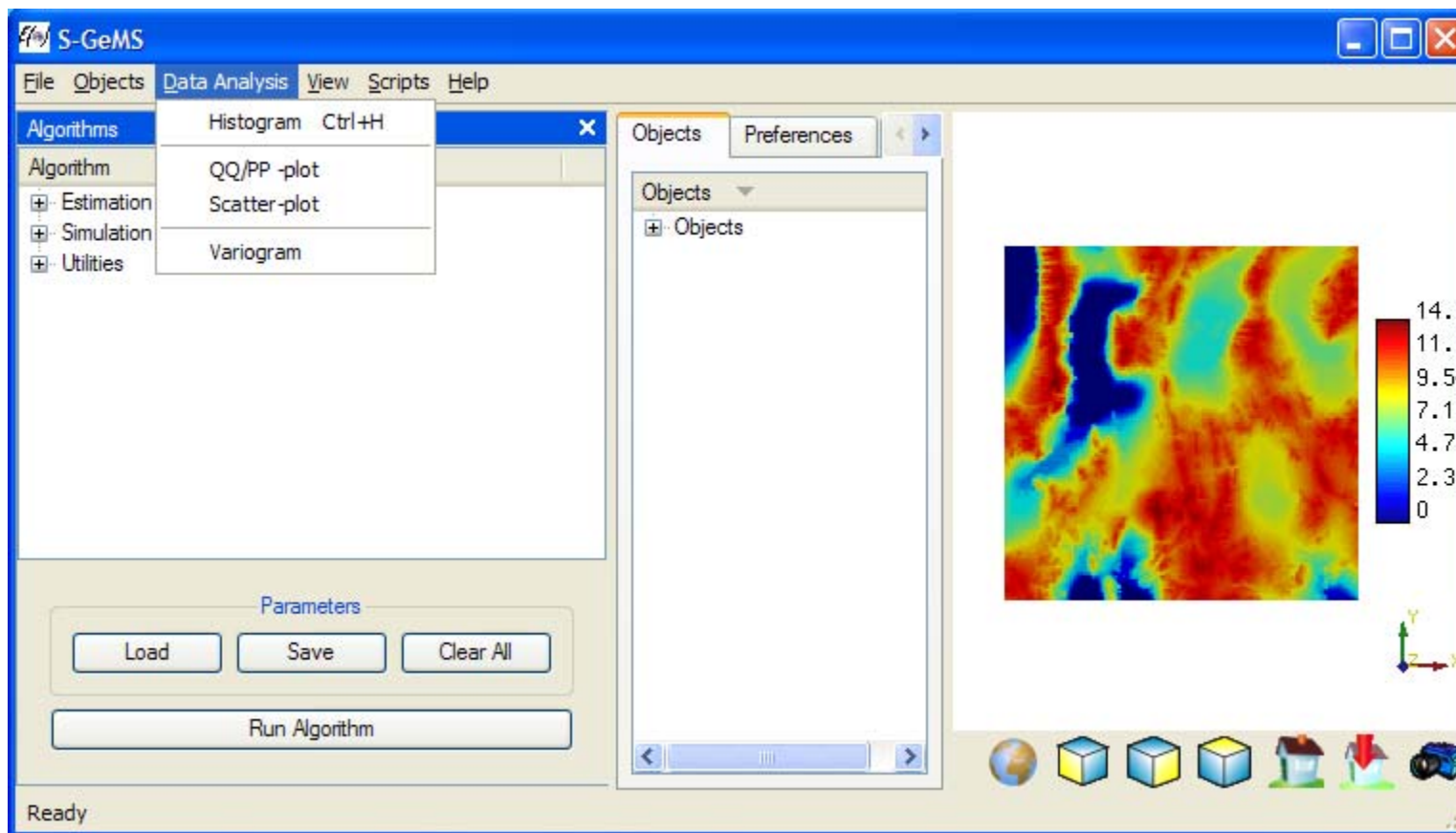
First click on <Data Analysis> and then on <Variogram>.
See page 187 for more information about this dataset.

C. SELECT FOLDER AND ATTRIBUTE
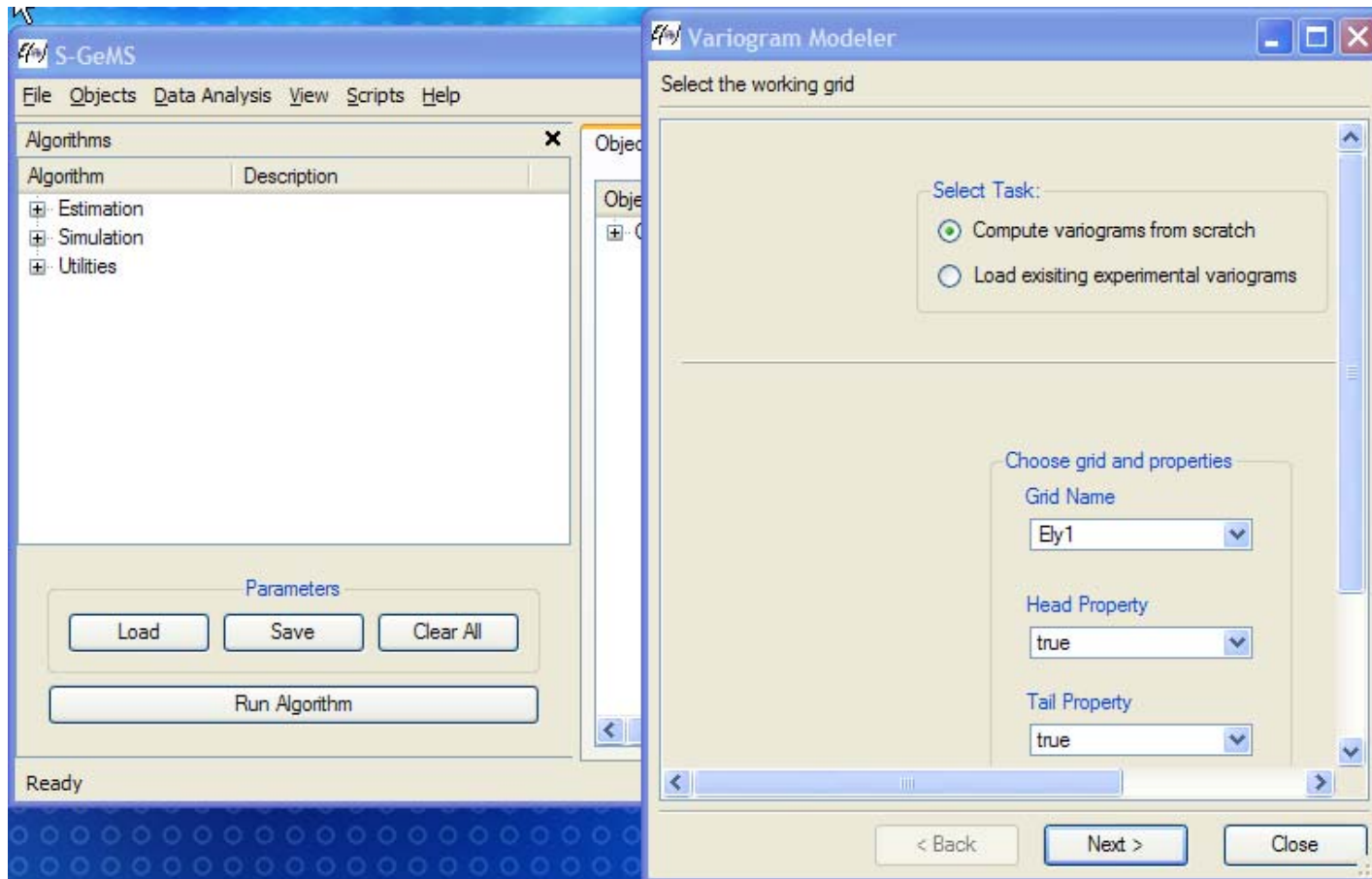
- Select <Head Property> equal to <Tail Property>.
- After making selections, click on <Next>.

D. DIRECTIONS AND LAGS

For grids, the azimuth is specified by directional cosines. For example, north in a horizontal place is (0, 1, 0).
For datasets, azimuth is specified as clockwise degrees from the north; so, for example, NE is 45º.

Upon specifying lags and directions, click on <Next>.

E – F. TREND DETECTION AND MODELING

- Hide the main menu screen;

- expand sideways semivariogram menu;

- click on <Window>, and then click on <Tile>;

- pick best fitting model parameters by trial and error.
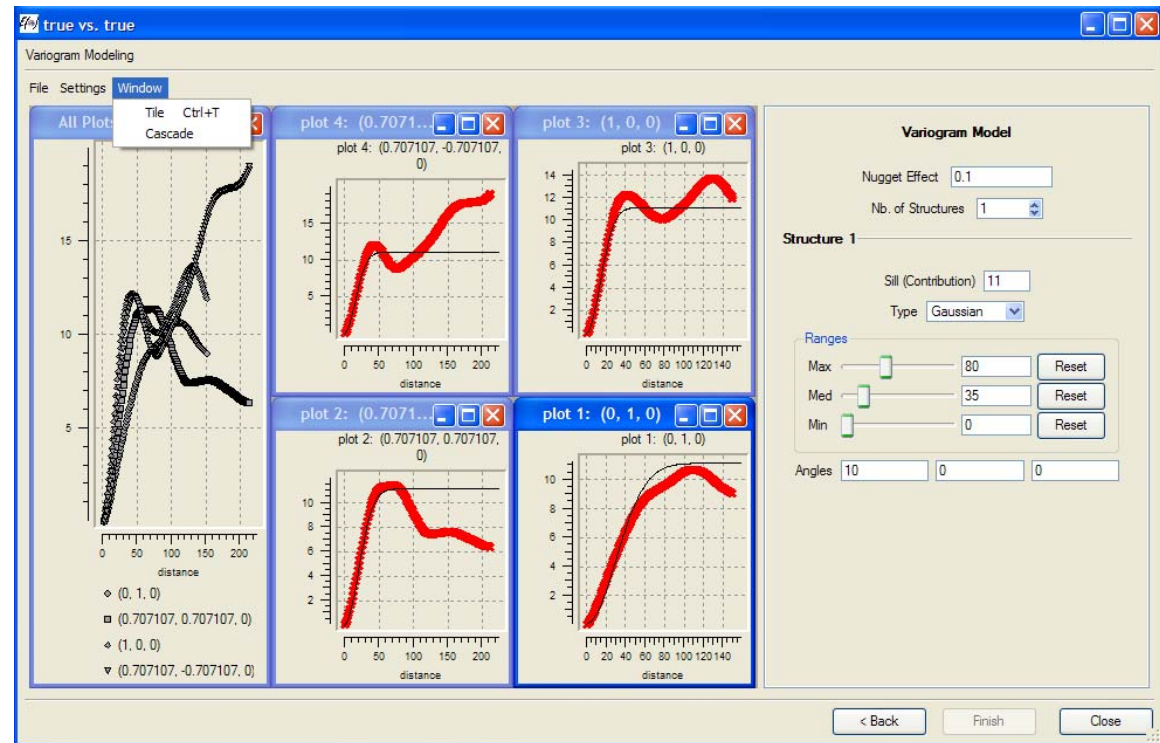
Here there is no trend because the semivariogram tends to stabilize at a value close to the sample variance. See pages 202–203 for modeling of a nonstationary dataset.

SAVING SEMIVARIOGRAMS

- Click on the <File> option of semivariogram modeling.
- Pick an option, for example, <Export Plot as Images>.
- Check <Show Grid> if it is desired to display it.
- Click on the three dots <…>.
- Select storage folder.
- Give generic name to semivariograms.
- Click on <Save> in both menus.
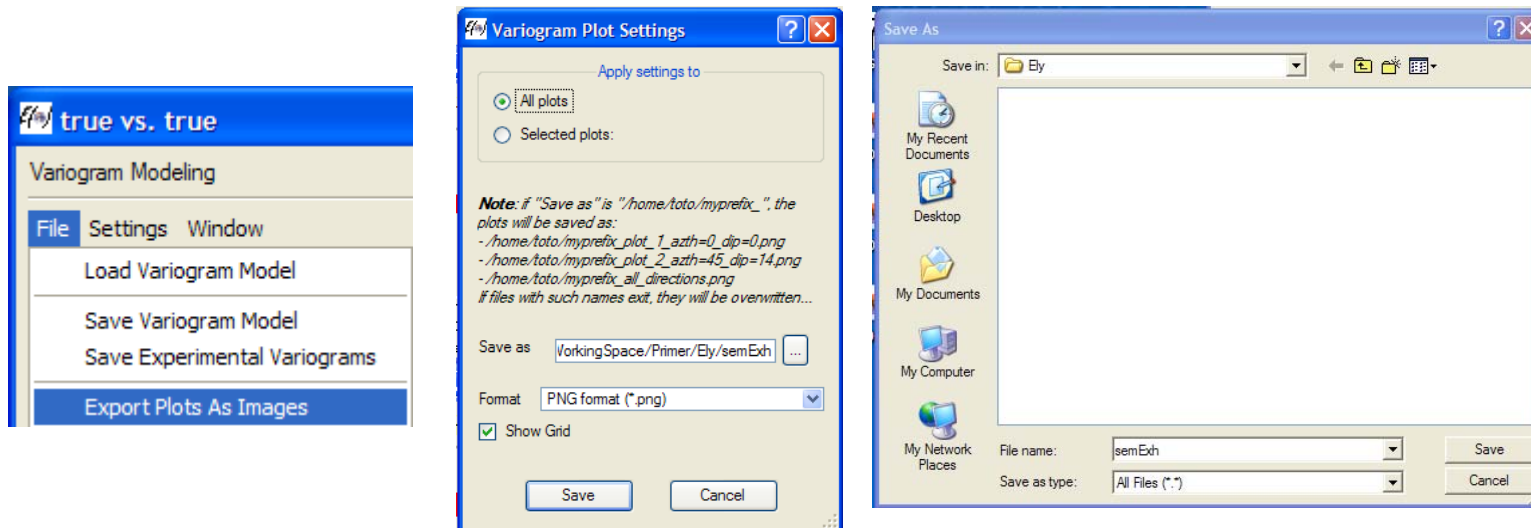
OTHER DIRECT APPLICATIONS

Direct applications of the semivariogram, not necessarily requiring to go through modeling, include:

- Detection of anisotropy by finding out if the semivariogram is different along different directions.
- Comparative studies of origins in the processes behind attributes based on the premise that different geneses lead to different semivariograms.
- Sampling design. In any direction, sampling interval should not exceed half the range.
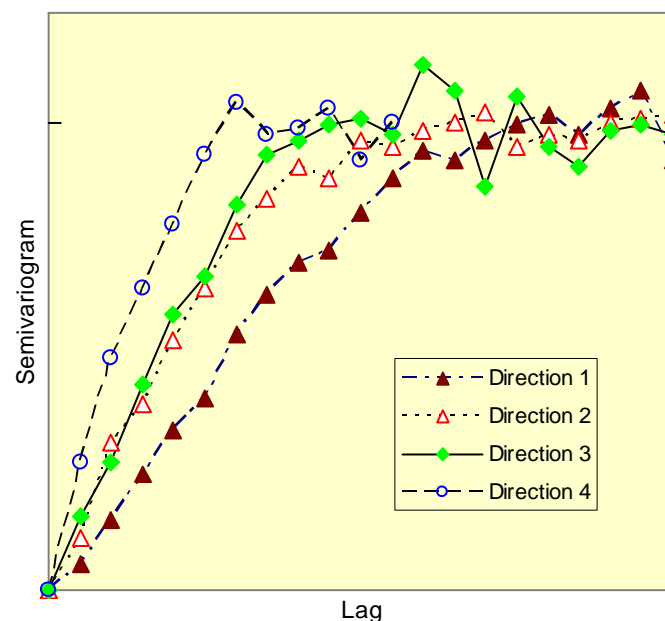
8. SIMPLE KRIGING

KRIGING

Kriging is a form of generalized linear regression for the formulation of an optimal spatial estimator in a minimum mean-square-error sense.

Given some observations at locations s_i, kriging provides an estimate and a standard error at locations, such as s_0, typically not considered in the sampling.

In contrast to classical linear regression, kriging takes into account observation volume and stochastic dependence among data.

The method works best inside the convex hull determined by the peripheral data.
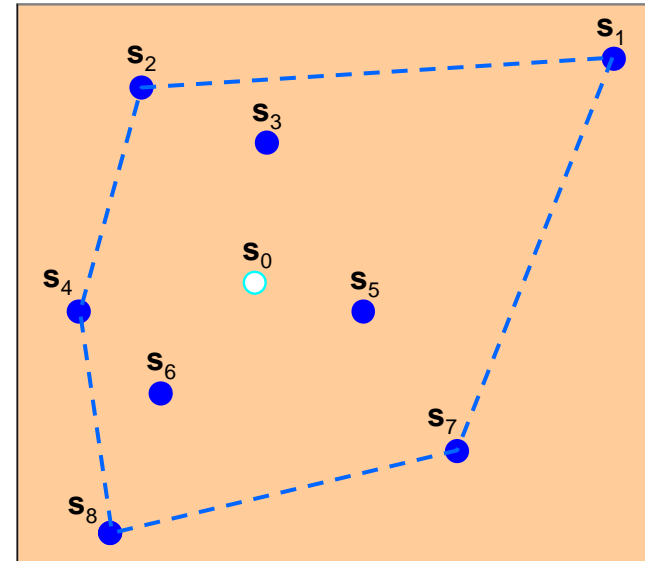
VARIANTS

There are several forms of kriging, each making different assumptions to adjust to different styles of attribute fluctuations.

- Simple kriging

- Ordinary kriging

- Universal kriging

- Lognormal kriging

- Indicator kriging

The last two types are simply any of the first three forms of kriging applied to transformed data as explained in chapter 6.

SIMPLE KRIGING

Simple kriging is the estimation of

$$z_{SK}^{*}(\mathbf{s}_0) = m + \sum_{i=1}^{n} \lambda_i \cdot (z(\mathbf{s}_i) - m),$$

where:

$z_{SK}^{*}(\mathbf{s}_0)$ is the estimator at geographical location $\mathbf{s}_0$;

$\mathbf{s}_i$ denotes the location of measurement i;

n is the number of observations to consider;

m is the mean of $Z(\mathbf{s})$;

λ_i is a real number, commonly referred to as weight.

NORMAL EQUATIONS

The originality of kriging is the use of weights, λ_i, such that they minimize the estimation error variance. Such weights turn out to be the solution of the following system of equations, called the normal system of equations:

$$\left. \begin{array}{rcl}
\displaystyle\sum_{i=1}^{k} \lambda_i \, \mathrm{Cov}\!\left(\mathbf{s}_i, \mathbf{s}_1\right) & = & \mathrm{Cov}\!\left(\mathbf{s}_0, \mathbf{s}_1\right) \\[2ex]
\displaystyle\sum_{i=1}^{k} \lambda_i \, \mathrm{Cov}\!\left(\mathbf{s}_i, \mathbf{s}_2\right) & = & \mathrm{Cov}\!\left(\mathbf{s}_0, \mathbf{s}_2\right) \\[1ex]
\cdots\cdots\cdots\cdots\cdots & \cdots & \cdots\cdots\cdots\cdots \\[1ex]
\displaystyle\sum_{i=1}^{k} \lambda_i \, \mathrm{Cov}\!\left(\mathbf{s}_i, \mathbf{s}_k\right) & = & \mathrm{Cov}\!\left(\mathbf{s}_0, \mathbf{s}_k\right)
\end{array} \right\}$$

DERIVATION OF NORMAL EQUATIONS

The steps are:

A. Make necessary assumptions.

B. Calculate error variance.

C. Express error variance in terms of covariances.

D. Find the weights that minimize the error variance.

E. Calculate the minimum error variance.

A. ASSUMPTIONS

1. The sample is a partial realization of a random function $Z(\mathbf{s})$, where $\mathbf{s}$ denotes spatial location.

2. $Z(\mathbf{s})$ is second order stationary, which implies:

$$\mathsf{E}[Z(\mathbf{s})] = m$$

$$\mathsf{E}[(Z(\mathbf{s}) - m)(Z(\mathbf{s} + \mathbf{h}) - m)] = Cov(\mathbf{h}),$$

   where $\mathsf{E}[\cdot]$ denotes expected value, m is a scalar constant, and $\mathbf{h}$ is a vectorial distance.

3. Unique to simple kriging is the assumption that the mean is not only constant but known.

   These assumptions make simple kriging the most restricted form of kriging.

   The weakness of this simplicity is the same as with any simple model: limitation in its applicability or suboptimal results if its use is forced.

B. ESTIMATION VARIANCE

By definition, the estimation variance is:

$$\sigma^2(\mathbf{s}_0) = \mathrm{Var}\left[Z^*_{SK}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right].$$

We know neither the random function $Z(\mathbf{s}_0)$ nor the estimator $Z^*_{SK}(\mathbf{s}_0)$, so we cannot directly calculate the variance of the difference.

The way to solve this conundrum is to transform the expression for the variance to have it in terms of quantities that we know.

C. ALGEBRAIC MANIPULATIONS (1)

The residual $Y(\mathbf{s})$ of a random function is the difference between the random function and its expected value. In this case:

$$Y(\mathbf{s}) = Z(\mathbf{s}) - \mathrm{E}[Z(\mathbf{s})] = Z(\mathbf{s}) - m$$

Expressing this estimation variance in terms of the residuals, we have:

$$\sigma^2(\mathbf{s}_0) = \mathrm{Var}\left[\sum_{i=1}^{k} \lambda_i Y(\mathbf{s}_i) - Y(\mathbf{s}_0)\right] = \mathrm{Var}\left[\sum_{i=0}^{k} \lambda_i Y(\mathbf{s}_i)\right]$$

if one defines $\lambda_0 = 1$.

ALGEBRAIC MANIPULATIONS (2)

From the properties of the variance (pages 31 and 52):

$$\sigma^2(\mathbf{s}_0) = \mathrm{E}\left[\left\{\sum_{i=0}^{k}\lambda_i Y(\mathbf{s}_i)\right\}^2\right] - \left\{\mathrm{E}\left[\sum_{i=0}^{k}\lambda_i Y(\mathbf{s}_i)\right]\right\}^2.$$

Expanding the squares and introducing the expectations inside the summations:

$$\sigma^2(\mathbf{s}_0) = \sum_{i=0}^{k}\sum_{j=0}^{k}\lambda_i\lambda_j\,\mathrm{E}\big[Y(\mathbf{s}_i)Y(\mathbf{s}_j)\big]$$

$$- \sum_{i=0}^{k}\sum_{j=0}^{k}\lambda_i\lambda_j\,\mathrm{E}\big[Y(\mathbf{s}_i)\big]\mathrm{E}\big[Y(\mathbf{s}_j)\big].$$

ALGEBRAIC MANIPULATIONS (3)

Grouping the terms by coefficients:

$$\sigma^2(\mathbf{s}_0) = \sum_{i=0}^{k} \sum_{j=0}^{k} \lambda_i \lambda_j \left\{ \mathrm{E}\left[Y(\mathbf{s}_i)Y(\mathbf{s}_j)\right] - \mathrm{E}\left[Y(\mathbf{s}_i)\right]\mathrm{E}\left[Y(\mathbf{s}_j)\right] \right\}.$$

Backtransforming to the random function and expanding:

$$\sigma^2(\mathbf{s}_0) = \sum_{i=0}^{k} \sum_{j=0}^{k} \lambda_i \lambda_j \left\{ \mathrm{E}\left[m^2 - mZ(\mathbf{s}_i) - mZ(\mathbf{s}_j) + Z(\mathbf{s}_i)Z(\mathbf{s}_j)\right] \right.$$

$$\left. - m^2 + m\mathrm{E}\left[Z(\mathbf{s}_i)\right] + m\mathrm{E}\left[Z(\mathbf{s}_j)\right] - \mathrm{E}\left[Z(\mathbf{s}_i)\right]\mathrm{E}\left[Z(\mathbf{s}_j)\right] \right\}.$$

Because of the cancelation of terms,

$$\sigma^2(\mathbf{s}_0) = \sum_{i=0}^{k} \sum_{j=0}^{k} \lambda_i \lambda_j \left\{ \mathrm{E}\left[Z(\mathbf{s}_i)Z(\mathbf{s}_j)\right] - \mathrm{E}\left[Z(\mathbf{s}_i)\right]\mathrm{E}\left[Z(\mathbf{s}_j)\right] \right\},$$

ALGEBRAIC MANIPULATIONS (4)

which is the same as:

$$\sigma^2(\mathbf{s}_0) = \sum_{i=0}^{k}\sum_{j=0}^{k} \lambda_i \lambda_j \, \text{Cov}(\mathbf{s}_i, \mathbf{s}_j).$$

Finally, remembering that $\lambda_0 = 1$:

$$\sigma^2(\mathbf{s}_0) = \text{Cov}(\mathbf{s}_0, \mathbf{s}_0) - 2\sum_{i=1}^{k} \lambda_i \, \text{Cov}(\mathbf{s}_0, \mathbf{s}_i)$$
$$+ \sum_{i=1}^{k}\sum_{j=1}^{k} \lambda_i \lambda_j \, \text{Cov}(\mathbf{s}_i, \mathbf{s}_j),$$

which is an expression in terms of covariance, a measurable property, hence a useful expression.

D. OPTIMAL WEIGHTS (1)

The expression for $\sigma^2(\mathbf{s}_0)$ is valid for any weights. Among the infinite weight combinations, we are interested in those that minimize the estimation variance associated with the estimator $Z^*_{SK}(\mathbf{s}_0)$.

The expression for $\sigma^2(\mathbf{s}_0)$ is quadratic in the unknowns, the weights. Under these circumstances, according to an operations research theorem, the necessary and sufficient condition to have a unique global nonnegative minimum is that the quadratic term must be larger than or equal to zero. This implies that the covariance function must be such that any combination, like the one in the general expression for $\sigma^2(\mathbf{s}_0)$, cannot be negative. In mathematics, such is the property of being positive definite. This is the condition a covariance model has to satisfy to be permissible.

OPTIMAL WEIGHTS (2)

The optimal weights are those that make zero all partial derivatives of $\sigma^2(\mathbf{s}_0)$ with respect to the weights:

$$\frac{\partial \sigma^2(\mathbf{s}_0)}{\partial \lambda_j} = -2\,\mathrm{Cov}(\mathbf{s}_0, \mathbf{s}_j) + 2\sum_{i=1}^{k} \lambda_i\,\mathrm{Cov}(\mathbf{s}_i, \mathbf{s}_j) = 0$$

for all locations j, $j = 1, 2, \ldots, k$.

Cancellation of factor 2 and rearrangement give:

$$\left.\begin{array}{rcl}
\displaystyle\sum_{i=1}^{k} \lambda_i\,\mathrm{Cov}(\mathbf{s}_i, \mathbf{s}_1) & = & \mathrm{Cov}(\mathbf{s}_0, \mathbf{s}_1) \\[2em]
\displaystyle\sum_{i=1}^{k} \lambda_i\,\mathrm{Cov}(\mathbf{s}_i, \mathbf{s}_2) & = & \mathrm{Cov}(\mathbf{s}_0, \mathbf{s}_2) \\[1em]
\cdots\cdots\cdots\cdots\cdots & \cdots & \cdots\cdots\cdots\cdots \\[1em]
\displaystyle\sum_{i=1}^{k} \lambda_i\,\mathrm{Cov}(\mathbf{s}_i, \mathbf{s}_k) & = & \mathrm{Cov}(\mathbf{s}_0, \mathbf{s}_k)
\end{array}\right\}$$

E. OPTIMAL ESTIMATION VARIANCE

Multiplying each optimality condition by λ_i and adding them together, we have a new optimality relationship:

$$\sum_{i=1}^{k} \lambda_i \sum_{j=1}^{k} \lambda_j \, \text{Cov}\big(\mathbf{s}_i, \mathbf{s}_j\big) = \sum_{i=1}^{k} \lambda_i \, \text{Cov}\big(\mathbf{s}_0, \mathbf{s}_i\big).$$

Introducing λ_i under the second summation and replacing the resulting double summation in the general expression for $\sigma^2\big(\mathbf{s}_0\big)$:

$$\sigma_{SK}^2\big(\mathbf{s}_0\big) = \text{Cov}\big(\mathbf{s}_0, \mathbf{s}_0\big) - \sum_{i=1}^{k} \lambda_i \, \text{Cov}\big(\mathbf{s}_0, \mathbf{s}_i\big).$$

SUMMARY

Simple kriging normal equations for optimal weights λ_i:

$$\left.\begin{array}{ccl}
\displaystyle\sum_{i=1}^{k} \lambda_i \, \mathrm{Cov}(\mathbf{s}_i, \mathbf{s}_1) & = & \mathrm{Cov}(\mathbf{s}_0, \mathbf{s}_1) \\
\displaystyle\sum_{i=1}^{k} \lambda_i \, \mathrm{Cov}(\mathbf{s}_i, \mathbf{s}_2) & = & \mathrm{Cov}(\mathbf{s}_0, \mathbf{s}_2) \\
\cdots\cdots\cdots\cdots\cdots & \cdots & \cdots\cdots\cdots\cdots \\
\displaystyle\sum_{i=1}^{k} \lambda_i \, \mathrm{Cov}(\mathbf{s}_i, \mathbf{s}_k) & = & \mathrm{Cov}(\mathbf{s}_0, \mathbf{s}_k)
\end{array}\right\}$$

Simple kriging estimation:

$$z_{SK}^{*}(\mathbf{s}_0) = m + \sum_{i=1}^{n} \lambda_i \left(z(\mathbf{s}_i) - m \right)$$

Simple kriging mean-square-error variance:

$$\sigma_{SK}^{2}(\mathbf{s}_0) = \mathrm{Cov}(\mathbf{s}_0, \mathbf{s}_0) - \sum_{i=1}^{k} \lambda_i \, \mathrm{Cov}(\mathbf{s}_0, \mathbf{s}_i)$$

MATRICES

A matrix is a rectangular array of numbers, such as **A**. When $n = m$, **A** is a square matrix of order n.

Transposing all rows and columns in **A** is denoted as $\mathbf{A}^T$.

Matrices are a convenient notation heavily used in dealing with large systems of linear equations, which notably reduce in size to just **A X** = **B** or **X** = $\mathbf{A}^{-1}\mathbf{B}$, where $\mathbf{A}^{-1}$ denotes the inverse of matrix **A**.

The main diagonal of a square matrix is the sequence of elements $a_{11}, a_{22}, \ldots, a_{nn}$ from upper left to lower right.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} b_1 & b_2 & \cdots & b_m \end{bmatrix}^T$$

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix}^T$$

SIMPLE KRIGING MATRIX FORMULATION

$$\begin{bmatrix} \mathrm{Cov}(\mathbf{s}_1,\mathbf{s}_1) & \cdots & \mathrm{Cov}(\mathbf{s}_k,\mathbf{s}_1) \\ \vdots & \cdots & \vdots \\ \mathrm{Cov}(\mathbf{s}_1,\mathbf{s}_k) & \ldots & \mathrm{Cov}(\mathbf{s}_k,\mathbf{s}_k) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{bmatrix} = \begin{bmatrix} \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_1) \\ \vdots \\ \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_K) \end{bmatrix}$$

$$Z_{SK}^{*}(\mathbf{s}_0) = m + \begin{bmatrix} Z(\mathbf{s}_1)-m \\ \vdots \\ Z(\mathbf{s}_k)-m \end{bmatrix}^{T} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{bmatrix}$$

$$\sigma_{SK}^{2}(\mathbf{s}_0) = \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_0) - \begin{bmatrix} \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_1) \\ \vdots \\ \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_k) \end{bmatrix}^{T} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{bmatrix}$$
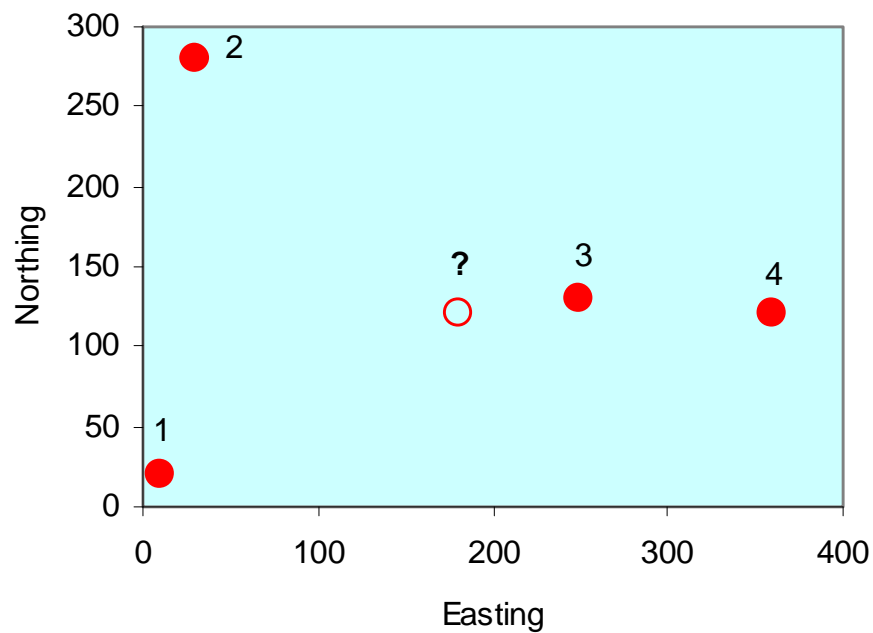
EXERCISE

SAMPLE

Index	Easting	Northing	Attribute
1	10	20	40
2	30	280	130
3	250	130	90
4	360	120	160
?	180	120	

m=110

$$\text{Cov}(\mathbf{h}) = 2000e^{-\mathbf{h}/250}$$

EXERCISE

For the previous configuration, using simple kriging:

A. Discuss the weight values.

B. Find the estimate.

C. Compute the estimation variance.

D. Calculate the estimation weights.

A. CALCULATION OF WEIGHTS (1)

Distance matrix among observations:

$$\begin{bmatrix} 0 & & & \\ 260.8 & 0 & & \\ 264.0 & 266.3 & 0 & \\ 364.0 & 366.7 & 110.4 & 0 \end{bmatrix}$$

Distance from estimation location to each observation:

$$\begin{bmatrix} 197.2 & 219.3 & 70.7 & 180.0 \end{bmatrix}^T$$

CALCULATION OF WEIGHTS (2)

Weights:

$$\begin{bmatrix} 2000 & & & \\ 704.8 & 2000 & & \\ 695.6 & 689.4 & 2000 & \\ 466.4 & 461.2 & 1285.8 & 2000 \end{bmatrix}^{-1} \begin{bmatrix} 908.7 \\ 831.8 \\ 1507.2 \\ 973.6 \end{bmatrix} = \begin{bmatrix} 0.185 \\ 0.128 \\ 0.646 \\ -0.001 \end{bmatrix}$$

B. REMARKS
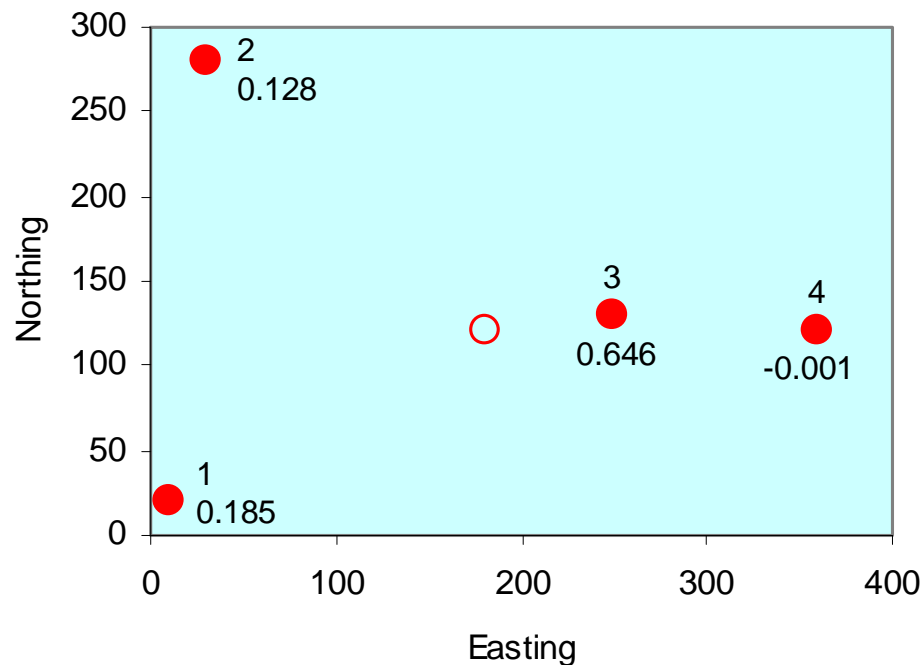
- Weight 4 is almost zero despite being closer to the estimation location than weight 1. This reduction known as screen effect is caused by the intermediate location of weight 3.

- Weight 4 is negative.

- Negative weights open the possibility to have estimates outside the extreme values of the data (nonconvexity).

C. ESTIMATE

$$z_{SK}^*\left(\mathbf{s}_0\right) = m + \begin{bmatrix} Z(\mathbf{s}_1) - m \\ \vdots \\ Z(\mathbf{s}_k) - m \end{bmatrix}^T \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{bmatrix}$$

$$z_{SK}^*\left(180,120\right) = 110 + \begin{bmatrix} 40 - 110 \\ 130 - 110 \\ 90 - 110 \\ 160 - 110 \end{bmatrix}^T \begin{bmatrix} 0.185 \\ 0.128 \\ 0.646 \\ -0.001 \end{bmatrix}$$

$$z_{SK}^*\left(180,120\right) = 86.7$$

D. ESTIMATION VARIANCE

$$\sigma_{SK}^2\left(\mathbf{s}_0\right) = \text{Cov}\left(\mathbf{s}_0, \mathbf{s}_0\right) - \begin{bmatrix} \text{Cov}\left(\mathbf{s}_0, \mathbf{s}_1\right) \\ \vdots \\ \text{Cov}\left(\mathbf{s}_0, \mathbf{s}_k\right) \end{bmatrix}^T \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{bmatrix}$$

$$\sigma_{SK}^2\left(180,120\right) = 2000 - \begin{bmatrix} 908.6 \\ 831.8 \\ 1507.2 \\ 973.6 \end{bmatrix}^T \begin{bmatrix} 0.185 \\ 0.128 \\ 0.646 \\ -0.001 \end{bmatrix}$$

$$\sigma_{SK}^2\left(180,120\right) = 752.9$$

PROPERTIES OF SIMPLE KRIGING

- Nonconvexity
- Screen effect
- Declustering
- Global unbiasedness
- Minimum mean-square-error optimality
- Exact interpolator
- When interpolation is exact, the error variance is 0
- Intolerance to duplicated sites
- The estimate is orthogonal to its error
- Independence to translation of Cartesian system

PROPERTIES OF SIMPLE KRIGING

- Independence of estimator to covariance scaling

- Scaling of covariance changes the estimation variance by same factor

- Estimation variance depends on data configuration

- Estimation variance does not depend directly on individual sample values

- If the weights sum up to 1, the estimator is independent of the value of the mean

MULTIGAUSSIAN KRIGING

- This is kriging of normal scores.
- The univariate marginal distribution of a multivariate normal distribution is normal, yet the converse is not true.
- Under multivariate normality, simple kriging is the best of all possible estimators, linear or nonlinear, biased or unbiased.
- Multivariate normality cannot be verified; at most, it cannot be rejected.
- Considering that the mean of normal scores is zero, if they are second order stationary, they provide a rare opportunity to apply simple kriging.
- Backtransformation of estimated normal scores outside the interval of variation for the normal scores of the data is highly uncertain.

9. ORDINARY KRIGING

BRILLIANT TRICK

One can rewrite the simple kriging estimator as

$$Z^*_{SK}(\mathbf{s}_0) = m\left(1 - \sum_{i=1}^{n} \lambda_i\right) + \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i).$$

If the weights sum up to 1, then the estimator is independent of *m*, thus resulting in a more general formulation applicable to samples with constant, but unknown mean (no trend).

It can be proved than the constraint has the additional benefit of making the estimator unbiased.

Ordinary kriging was a first formulation of improved simple kriging. Simple kriging roots predate geostatistics. We will see that one can still find weights minimizing the mean square estimation error.

ORDINARY KRIGING

The ordinary kriging estimator is:

$$Z_{OK}^{*}(\mathbf{s}_0) = \begin{cases} \sum_{i=1}^{k} \lambda_i Z(\mathbf{s}_i) \\ \sum_{i=1}^{k} \lambda_i = 1, \end{cases}$$

where:
$z_{OK}^{*}(\mathbf{s}_0)$ denotes estimation at geographic location $\mathbf{s}_0$;
$\mathbf{s}_i$ is the location of measurement i;
k denotes the number of observations to consider;
m is the mean of $Z(\mathbf{s})$; and
λ_i is a real weight.

DERIVATION OF EQUATIONS

The whole purpose of ordinary kriging is to find optimal weights that minimize the mean square estimation error. The steps are the same as those required to formulate simple kriging:

A. Make necessary assumptions;

B. Find expression for estimation error variance;

C. Modify estimation error variance equation to have it in terms of covariances;

D. Out of all possible weights, find those that minimize the error variance.

E. Replace the optimal weights in the expression for error variance. This is the kriging variance.

A. STRONGER ASSUMPTIONS

1. The sample is a partial realization of a random function $Z(\mathbf{s})$, where $\mathbf{s}$ denotes spatial location.

2. $Z(\mathbf{s})$ is second order stationary, that is:

$$\mathrm{E}[Z(\mathbf{s})] = m$$

$$\mathrm{E}[(Z(\mathbf{s}) - m)(Z(\mathbf{s} + \mathbf{h}) - m)] = \mathrm{Cov}(\mathbf{h}),$$

where $\mathrm{E}[\cdot]$ denotes expected value, m is a scalar constant, and $\mathbf{h}$ is a vectorial distance.

As in simple kriging, the mean is still constant, but now it can be unknown.

WEAKER ASSUMPTIONS

The following is a set of slightly weaker assumptions, in the sense that they do not require that the variance be finite.

1. The sample is a partial realization of a random function $Z(\mathbf{s})$, where $\mathbf{s}$ denotes spatial location.

2. $Z(\mathbf{s})$ honors the intrinsic hypothesis, meaning that:

$$\mathsf{E}[Z(\mathbf{s})] = m$$
$$\mathsf{Var}[Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h})] = 2\gamma(\mathbf{h}),$$

where $\gamma(\mathbf{h})$ is the semivariogram.

The mean must be constant, but because it will not enter into the calculations, it can be unknown.

B. ESTIMATION VARIANCE

The estimation variance is the error variance:

$$\sigma^2(\mathbf{s}_0) = \text{Var}\left[Z^*_{OK}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right],$$

where $Z(\mathbf{s}_0)$ is the random function at the estimation location $\mathbf{s}_0$ and $Z^*_{OK}(\mathbf{s}_0)$ is the ordinary kriging estimator.

   We know neither of these, thus it is not possible to calculate $\sigma^2(\mathbf{s}_0)$. The way out is to transform the expression for the variance to have it in terms of quantities that we can calculate. This will require even more laborious algebraic manipulations than for simple kriging. Nothing comes for free!

LEMMA 1 (1)

Let $Z(\mathbf{s}_i)$ be a random variable of a continuous random function. Then for any coefficient λ_i :

$$\mathrm{E}\left[\left\{\sum_{i=1}^{k}\lambda_i Z(\mathbf{s}_i)\right\}^2\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_i\lambda_j\,\mathrm{E}\left[Z(\mathbf{s}_i)Z(\mathbf{s}_j)\right].$$

Proof:

One can regard $\lambda_i\lambda_j$ as a new constant and $Z(\mathbf{s}_i)Z(\mathbf{s}_j)$ as a new variable.

LEMMA 1 (2)

Then, by the distributive and commutative properties of the expectation:

$$E\left[\left\{\sum_{i=1}^{k}\lambda_i Z(\mathbf{s}_i)\right\}^2\right] = \lambda_1\lambda_i\sum_{i=1}^{k}E[Z(\mathbf{s}_1)Z(\mathbf{s}_i)]$$

$$+\lambda_2\lambda_i\sum_{i=1}^{k}E[Z(\mathbf{s}_2)Z(\mathbf{s}_i)]+\ldots+\lambda_k\lambda_i\sum_{i=1}^{k}E[Z(\mathbf{s}_k)Z(\mathbf{s}_i)]$$

and the proof follows by condensing the expression using a second summation index.

LEMMA 2 (1)

Let $Z(\mathbf{s}_i)$ be a random variable of a continuous random function. Then for any coefficient λ_i:

$$\mathrm{Var}\left[\left\{\sum_{i=1}^{k}\lambda_i Z(\mathbf{s}_i)\right\}^2\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_i\lambda_j\,\mathrm{Cov}\left[\mathbf{s}_i,\mathbf{s}_j\right].$$

Proof:

By the definition of variance (pages 31 and 52):

$$\mathrm{Var}\left[\sum_{i=i}^{k}\lambda_i Z(\mathbf{s}_i)\right] = \mathrm{E}\left[\left\{\sum_{i=1}^{k}\lambda_i Z(\mathbf{s}_i)\right\}^2\right] - \left\{\mathrm{E}\left[\sum_{i=1}^{k}\lambda_i Z(\mathbf{s}_i)\right]\right\}^2.$$

LEMMA 2 (2)

By Lemma 1:

$$\text{Var}\left[\sum_{i=1}^{k}\lambda_i Z(\mathbf{s}_i)\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_i\lambda_j\, \text{E}\big[Z(\mathbf{s}_i)Z(\mathbf{s}_j)\big]$$

$$-\sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_i\lambda_j\, \text{E}\big[Z(\mathbf{s}_i)\big]\text{E}\big[Z(\mathbf{s}_j)\big].$$

Factoring,

$$\text{Var}\left[\sum_{i=1}^{k}\lambda_i Z(\mathbf{s}_i)\right] = \sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_i\lambda_j\Big\{\text{E}\big[Z(\mathbf{s}_i)Z(\mathbf{s}_j)\big]$$

$$-\text{E}\big[Z(\mathbf{s}_i)\big]\text{E}\big[Z(\mathbf{s}_j)\big]\Big\}.$$

The fact that the difference is equal to the covariance of $Z(\mathbf{s}_i)$ and $Z(\mathbf{s}_j)$ proves the lemma.

LEMMA 3 (1)

Let $Z(\mathbf{s})$ be an intrinsic random function. Then:

$$\gamma(\mathbf{s}_i - \mathbf{s}_j) = \gamma(\mathbf{s}_i - \mathbf{s}) + \gamma(\mathbf{s}_j - \mathbf{s})$$
$$- \mathrm{Cov}(\mathbf{s}_i - \mathbf{s}, \mathbf{s}_j - \mathbf{s}).$$

Proof:

By the stationarity assumption:

$$\gamma(\mathbf{s}_i - \mathbf{s}_j) = \frac{1}{2} \mathrm{E}\left[\{Z(\mathbf{s}_i) - Z(\mathbf{s}_j)\}^2\right],$$

which does not change by adding and subtracting a third variate

$$\gamma(\mathbf{s}_i - \mathbf{s}_j) = \frac{1}{2} \mathrm{E}\left[\{\{Z(\mathbf{s}_i) - Z(\mathbf{s})\} - \{Z(\mathbf{s}_j) - Z(\mathbf{s})\}\}^2\right].$$

LEMMA 3 (2)

Expanding:

$$\gamma\left(\mathbf{s}_i - \mathbf{s}_j\right) = \frac{1}{2}\mathsf{E}\left[\{Z(\mathbf{s}_i) - Z(\mathbf{s})\}^2\right] + \frac{1}{2}\mathsf{E}\left[\{Z(\mathbf{s}_i) - Z(\mathbf{s})\}^2\right]$$
$$- \mathsf{E}\left[Z(\mathbf{s}_i) - Z(\mathbf{s}), Z(\mathbf{s}_j) - Z(\mathbf{s})\right],$$

and again, by the stationary assumption:

$$\gamma\left(\mathbf{s}_i - \mathbf{s}_j\right) = \gamma\left(\mathbf{s}_i - \mathbf{s}\right) + \gamma\left(\mathbf{s}_j - \mathbf{s}\right)$$
$$- \mathsf{E}\left[Z(\mathbf{s}_i) - Z(\mathbf{s}), Z(\mathbf{s}_j) - Z(\mathbf{s})\right].$$

The proof follows from the definition of the covariance because of the special fact that the mean of the differences is zero because the mean is constant.

C. NEW ERROR VARIANCE EXPRESSION (1)

Replacing the estimator by its definition in the expression for the error variance:

$$\sigma^2(\mathbf{s}_0) = \text{Var}\left[Z_{OK}^*(\mathbf{s}_0) - Z(\mathbf{s}_0)\right] = \text{Var}\left[\sum_{i=1}^{k} \lambda_i (Z(\mathbf{s}_i) - Z(\mathbf{s}_0))\right]$$

because $\sum_{i=1}^{k} \lambda_i = 1$. From Lemma 2,

$$\sigma^2(\mathbf{s}_0) = \sum_{i=1}^{k}\sum_{j=1}^{k} \lambda_i \lambda_j \, \text{Cov}(\mathbf{s}_i - \mathbf{s}_0, \mathbf{s}_j - \mathbf{s}_0),$$

which, by Lemma 3, is:

$$\sigma^2(\mathbf{s}_0) = \sum_{i=1}^{k}\sum_{j=1}^{k} \lambda_i \lambda_j \left\{\gamma(\mathbf{s}_i - \mathbf{s}_0) + \gamma(\mathbf{s}_j - \mathbf{s}_0) - \gamma(\mathbf{s}_i - \mathbf{s}_j)\right\}.$$

NEW ERROR VARIANCE EXPRESSION (2)

Expanding, because of the constraint $\displaystyle\sum_{i=1}^{k} \lambda_i = 1$:

$$\sigma^2(\mathbf{s}_0) = \sum_{i=1}^{k} \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_0) + \sum_{j=1}^{k} \lambda_j \gamma(\mathbf{s}_j - \mathbf{s}_0)$$

$$- \sum_{i=1}^{k} \sum_{j=1}^{k} \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j).$$

The first two terms in the right-hand side of the equation are the same, thus:

$$\sigma^2(\mathbf{s}_0) = 2\sum_{i=1}^{k} \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_0) - \sum_{i=1}^{k} \sum_{j=1}^{k} \lambda_i \lambda_j \gamma(\mathbf{s}_i - \mathbf{s}_j).$$

169

D. OPTIMAL WEIGHTS

The expression for $\sigma^2(\mathbf{s}_0)$ is quadratic in the unknowns, the weights. In such case, the necessary and sufficient condition to have a unique global minimum is that the quadratic term must be equal to or larger than zero, which implies that the semivariogram must be negative definite. Thus we need to employ only permissible models.

The expression for $\sigma^2(\mathbf{s}_0)$ is valid for any weights. Among the infinite weight combinations, we are interested in those minimizing the estimation variance associated with the estimator $Z^*_{OK}(\mathbf{s}_0)$. This is a constrained operations research problem, best solved by the Lagrange method of multipliers.

LAGRANGE METHOD

The Lagrange method of multipliers is a procedure for the optimization of a function of several variables $f(y_1, y_2, \ldots, y_n)$ constrained by $\phi_1, \phi_2, \ldots, \phi_m$.

The solution is found by equating to 0 the *n* partial derivatives of an unconstrained auxiliary function

$$f(y_1, y_2, \ldots, y_n) + \sum_{i=1}^{m} \mu_i \phi_i$$

with respect to all y_i, regarding the Lagrange multipliers, μ_i, as constants.

The resulting *n* equations plus the original *m* constraints provide a system of equations from which the unknown y_i and the auxiliary variables, μ_i, can be calculated.

TRIVIAL LAGRANGE EXAMPLE

To find the maximum area of a rectangle whose perimeter is the given value p, it is necessary to find the maximum value of the product of the sides $a \cdot b$, subject to the constraint $2a + 2b - p = 0$.

The auxiliary function is in this case
$$u = ab + \mu(2a + 2b - p),$$

which after differentiation with respect to the unknowns a and b gives
$$\left.\begin{array}{r} a + 2\mu = 0 \\ b + 2\mu = 0 \\ 2a + 2b = p \end{array}\right\}$$

whose solution is $a = b = 0.25p$, $\mu = -0.125$.

NORMAL EQUATIONS

The Lagrange auxiliary function in this case is:

$$L = 2\sum_{j=1}^{k} \lambda_j \gamma(\mathbf{s}_j, \mathbf{s}_0) - \sum_{i=1}^{k}\sum_{j=1}^{k} \lambda_i \lambda_j \gamma(\mathbf{s}_i, \mathbf{s}_j) + 2\mu\left(\sum_{j=1}^{k} \lambda_j - 1\right).$$

The derivatives with respect to the unknowns are:

$$\frac{\partial L}{\partial \lambda_j} = 2\gamma(\mathbf{s}_j - \mathbf{s}_0) - \sum_{i=1}^{k} \lambda_i \gamma(\mathbf{s}_i - \mathbf{s}_j) + 2\mu = 0, \text{ for } j = 1, \ldots, k.$$

Thus

$$\left.\begin{array}{c} \sum_{i=1}^{k} \lambda_i \gamma(\mathbf{s}_1, \mathbf{s}_i) - \mu = \gamma(\mathbf{s}_1, \mathbf{s}_0) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \sum_{i=1}^{k} \lambda_i \gamma(\mathbf{s}_k, \mathbf{s}_i) - \mu = \gamma(\mathbf{s}_k, \mathbf{s}_0) \\ \sum_{i=1}^{k} \lambda_i = 1 \end{array}\right\}$$

E. OPTIMAL ESTIMATION VARIANCE

Multiplying each optimality condition by λ_i and adding them together, we have the new optimality relationship:

$$\sum_{i=1}^{k} \lambda_i \sum_{j=1}^{k} \lambda_j \gamma\left(\mathbf{s}_i, \mathbf{s}_j\right) = \sum_{i=1}^{k} \lambda_i \gamma\left(\mathbf{s}_i, \mathbf{s}_0\right) + \mu \sum_{i=1}^{k} \lambda_i .$$

Because the weights sum up to 1, introducing λ_i under the second summation and replacing the resulting double summation in the general expression for the estimation variance, one obtains:

$$\sigma_{OK}^2\left(\mathbf{s}_0\right) = 2 \sum_{i=1}^{k} \lambda_i \gamma\left(\mathbf{s}_i, \mathbf{s}_0\right) - \sum_{i=1}^{k} \lambda_i \gamma\left(\mathbf{s}_i, \mathbf{s}_0\right) - \mu$$

or

$$\sigma_{OK}^2\left(\mathbf{s}_0\right) = \sum_{i=1}^{k} \lambda_i \gamma\left(\mathbf{s}_i, \mathbf{s}_0\right) - \mu .$$

SUMMARY

Ordinary kriging normal equations for optimal weights λ_i :

$$\left.\begin{array}{c} \displaystyle\sum_{i=1}^{k}\lambda_i\gamma(\mathbf{s}_1,\mathbf{s}_i)-\mu=\gamma(\mathbf{s}_0,\mathbf{s}_1) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \displaystyle\sum_{i=1}^{k}\lambda_i\gamma(\mathbf{s}_k,\mathbf{s}_i)-\mu=\gamma(\mathbf{s}_0,\mathbf{s}_k) \\ \displaystyle\sum_{i=1}^{k}\lambda_i=1 \end{array}\right\}$$

Ordinary kriging estimate

$$z_{OK}^{*}(\mathbf{s}_0)=\sum_{i=1}^{n}\lambda_i z(\mathbf{s}_i)$$

Ordinary kriging mean-square-error variance:

$$\sigma_{OK}^{2}(\mathbf{s}_0)=\sum_{i=1}^{k}\lambda_i\gamma(\mathbf{s}_i,\mathbf{s}_0)-\mu$$

INTRINSIC MATRIX FORMULATION

$$\begin{bmatrix} \gamma(\mathbf{s}_1,\mathbf{s}_1) & \cdots & \gamma(\mathbf{s}_k,\mathbf{s}_1) & 1 \\ \vdots & \cdots & \vdots & 1 \\ \gamma(\mathbf{s}_1,\mathbf{s}_k) & \cdots & \gamma(\mathbf{s}_k,\mathbf{s}_k) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ -\mu \end{bmatrix} = \begin{bmatrix} \gamma(\mathbf{s}_0,\mathbf{s}_1) \\ \vdots \\ \gamma(\mathbf{s}_0,\mathbf{s}_K) \\ 1 \end{bmatrix}$$

$$Z_{OK}^{*}(\mathbf{s}_0) = \begin{bmatrix} Z(\mathbf{s}_1) \\ \vdots \\ Z(\mathbf{s}_k) \\ 0 \end{bmatrix}^{T} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ -\mu \end{bmatrix}$$

$$\sigma_{OK}^{2}(s_0) = \begin{bmatrix} \gamma(\mathbf{s}_0,\mathbf{s}_1) \\ \vdots \\ \gamma(\mathbf{s}_0,\mathbf{s}_k) \\ 1 \end{bmatrix}^{T} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ -\mu \end{bmatrix}$$

STATIONARY MATRIX FORMULATION

$$\begin{bmatrix} \mathrm{Cov}(\mathbf{s}_1,\mathbf{s}_1) & \cdots & \mathrm{Cov}(\mathbf{s}_k,\mathbf{s}_1) & 1 \\ \vdots & \cdots & \vdots & 1 \\ \mathrm{Cov}(\mathbf{s}_1,\mathbf{s}_k) & \cdots & \mathrm{Cov}(\mathbf{s}_k,\mathbf{s}_k) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \mu \end{bmatrix} = \begin{bmatrix} \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_1) \\ \vdots \\ \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_K) \\ 1 \end{bmatrix}$$

$$Z^*_{OK}(\mathbf{s}_0) = \begin{bmatrix} Z(\mathbf{s}_1) \\ \vdots \\ Z(\mathbf{s}_k) \\ 0 \end{bmatrix}^T \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \mu \end{bmatrix}$$

$$\sigma^2_{OK}(\mathbf{s}_0) = \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_0) - \begin{bmatrix} \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_1) \\ \vdots \\ \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_k) \\ 1 \end{bmatrix}^T \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \mu \end{bmatrix}$$

EXERCISE

Index	Easting	Northing	Attribute
1	10	20	40
2	30	280	130
3	250	130	90
4	360	120	160
?	180	120	

SAMPLE

$$\mathrm{Cov}(\mathbf{h}) = 2000 e^{-\mathbf{h}/250}$$

EXERCISE

Apply ordinary kriging to the previous sample to do the following:

A.   Find the estimation weights.

B.   Note any interesting peculiarities in the weight values.

C.   Use the data and weights to evaluate the estimate.

D.   Calculate the estimation variance.

   Compare to the results obtained using simple kriging.

A. ORDINARY KRIGING WEIGHTS

Distance matrices:

$$\begin{bmatrix} 0 & & & \\ 260.8 & 0 & & \\ 264.0 & 266.3 & 0 & \\ 364.0 & 366.7 & 110.4 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 197.2 & 219.3 & 70.7 & 180.0 \end{bmatrix}^T$$

Weights and Lagrange multiplier:

$$\begin{bmatrix} 2000 & & & & \\ 704.8 & 2000 & & & \\ 695.6 & 689.4 & 2000 & & \\ 466.4 & 461.2 & 1285.8 & 2000 & \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 908.7 \\ 831.8 \\ 1507.2 \\ 973.6 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.185 \\ 0.128 \\ 0.646 \\ -0.001 \\ -42.714 \end{bmatrix}$$

USGS

B. WEIGHT COMPARISON
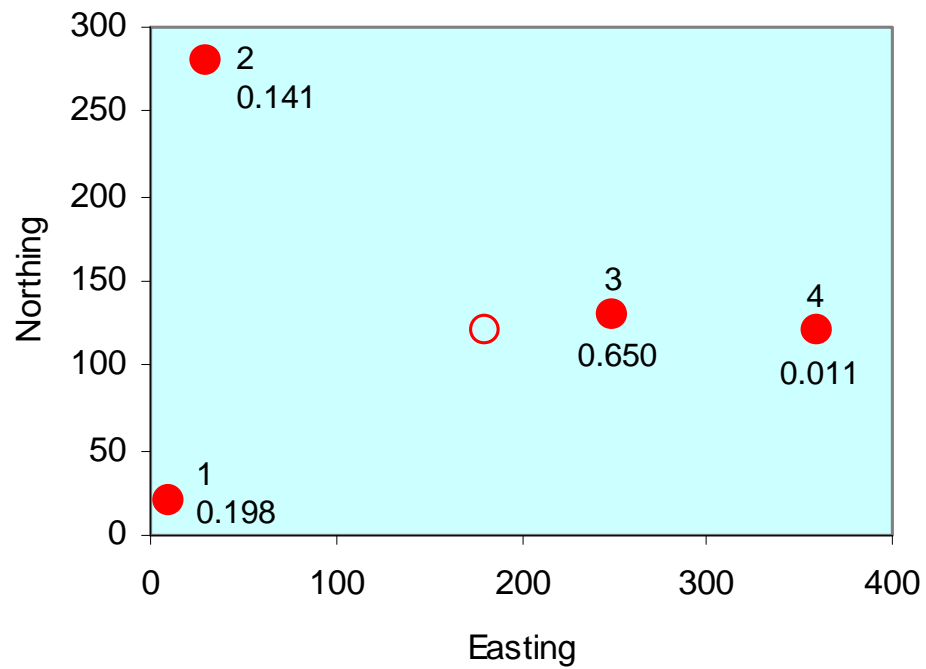
Weight	SK	OK
1	0.185	0.198
2	0.128	0.141
3	0.646	0.650
4	-0.001	0.011
Sum	0.958	1.000

C. ESTIMATES

Using the subscripts OK to denote ordinary kriging and SK for simple kriging,

$$z^*_{OK}(\mathbf{s}_0) = \begin{bmatrix} Z(\mathbf{s}_1) \\ \vdots \\ Z(\mathbf{s}_k) \\ 0 \end{bmatrix}^T \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ -\mu \end{bmatrix}$$

$$z^*_{OK}(180,120) = \begin{bmatrix} 40 \\ 130 \\ 90 \\ 160 \end{bmatrix}^T \begin{bmatrix} 0.198 \\ 0.141 \\ 0.650 \\ 0.011 \end{bmatrix}$$

$$z^*_{OK}(180,120) = 86.6$$

$$z^*_{SK}(180,120) \text{ was } 86.7$$

D. ESTIMATION VARIANCE

$$\sigma^2_{OK}(\mathbf{s}_0) = \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_0) - \begin{bmatrix} \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_1) \\ \vdots \\ \mathrm{Cov}(\mathbf{s}_0,\mathbf{s}_k) \\ 1 \end{bmatrix}^T \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \mu \end{bmatrix}$$

$$\sigma^2_{OK}(180,120) = 2000 - \begin{bmatrix} 908.6 \\ 831.8 \\ 1507.2 \\ 973.6 \\ 1 \end{bmatrix}^T \begin{bmatrix} 0.198 \\ 0.141 \\ 0.650 \\ 0.011 \\ -42.714 \end{bmatrix}$$

$$\sigma^2_{OK}(180,120) = 754.8$$

$\sigma^2_{SK}(180,120)$ was 752.9

PROPERTIES

Ordinary kriging shares the same properties with simple kriging. In addition,

- $\sigma_{OK}^2(\mathbf{s}_0) \geq \sigma_{SK}^2(\mathbf{s}_0)$

- For the trivial case of considering only one observation in the calculation of the estimate,

$$\sigma_{OK}^2(\mathbf{s}_0) = 2 \cdot \gamma(\mathbf{h}_{01}),$$

  where $\mathbf{h}_{01}$ is the distance from the observation, $z(\mathbf{s}_1)$, to the estimation location, $\mathbf{s}_0$.

SEARCH NEIGHBORHOOD

- Because of the screen effect and numerical instabilities, it is recommended that only the closest observations to the estimation location be used.

- Three observations is a reasonable bare minimum and 25 is more than adequate.

- Use octant search to further ensure good radial distribution.
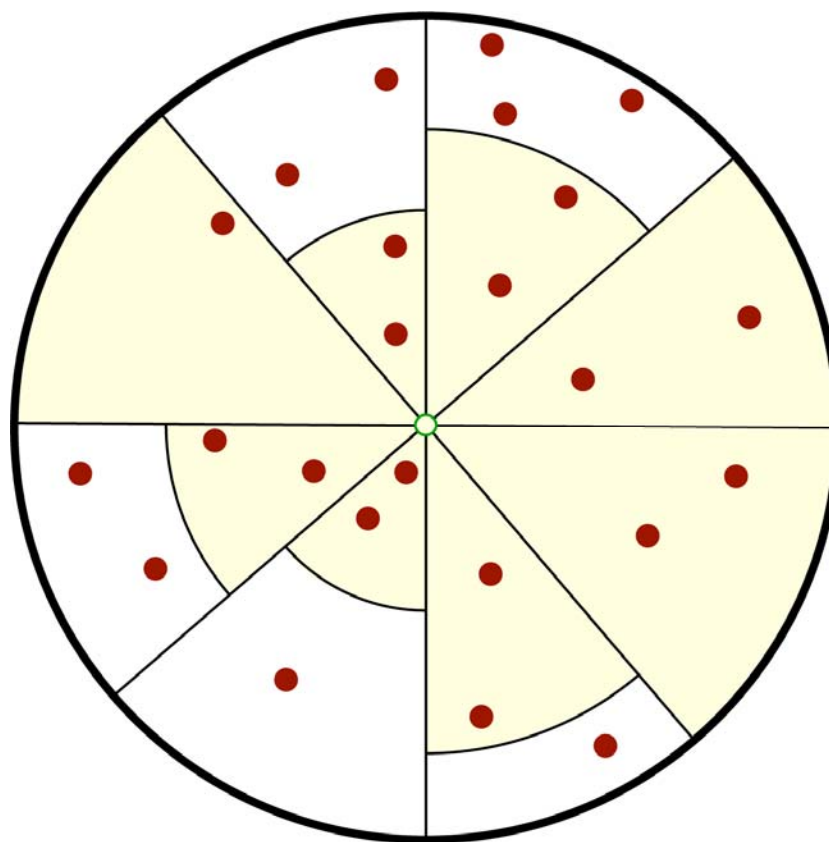
SUBSAMPLE SIZE SENSITIVITY FOR THE CASE OF THE UNCF DATASET

k	λ_1	λ_2	λ_3	λ_4	λ_5	$\sum_{i=6}^{k} \lambda_i$	z_{OK}^*	σ_{OK}^*
5	0.351	0.311	0.043	0.157	0.137	---	0.766	0.159
5	0.358	0.328	0.047	0.157	0.137	-0.047	0.750	0.159
5	0.356	0.325	0.050	0.178	0.165	-0.074	0.767	0.158
10	0.355	0.324	0.050	0.178	0.170	-0.077	0.756	0.158
15	0.355	0.324	0.051	0.178	0.169	-0.077	0.753	0.158
20	0.355	0.324	0.052	0.178	0.169	-0.078	0.754	0.158
25	0.355	0.324	0.052	0.178	0.169	-0.078	0.758	0.158

THE ELY WEST SAMPLE
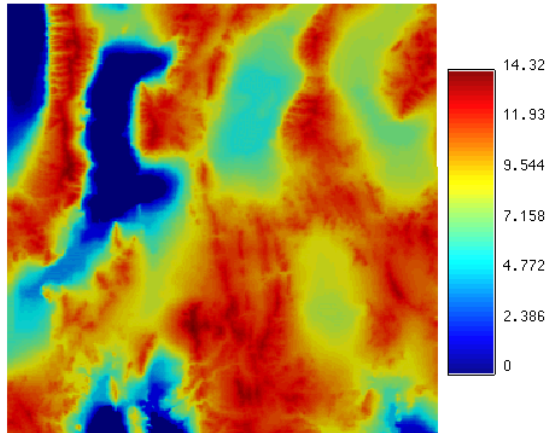
This is a semi-artificial, exhaustive dataset that will be used extensively to illustrate and compare several methods.

The data are transformed digital elevations in an area in Nevada with extreme cases of continuity: flat dry lakes surrounded by rugged sierras running roughly north-south.

The sampling is perfectly regular on a 286 by 286 grid, making 81,796 observations with a Gaussian anisotropic semivariogram (p. 124): $\gamma_E(\mathbf{h}) = 0.1 + 11 \cdot \text{Gaussian}(\text{N10E},80,34)$.

The exhaustive sample will be used solely for comparisons. All calculations will be done employing the same subset of size 40. Based partly on crossvalidation, its semivariogram is $\gamma(\mathbf{h}) = 0.1 + 13 \cdot \text{Gaussian}(\text{N15E},80,35)$.

ORDINARY KRIGING WITH SGeMS
A. LOAD SAMPLE
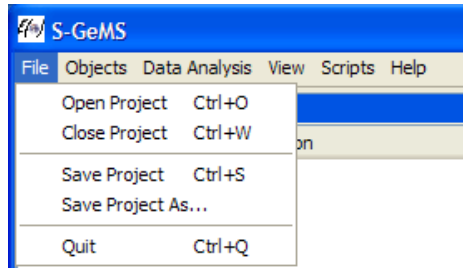
Load data according to instructions in chapter 4.

If data previously loaded and saved:

- click on <File>;

- click on <Open Project>;

- select folder and click <OK>.
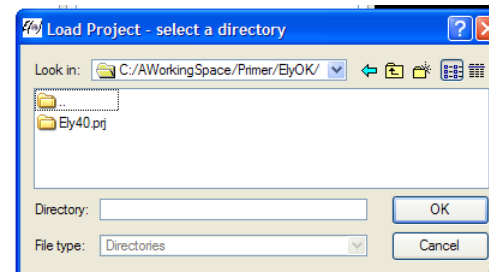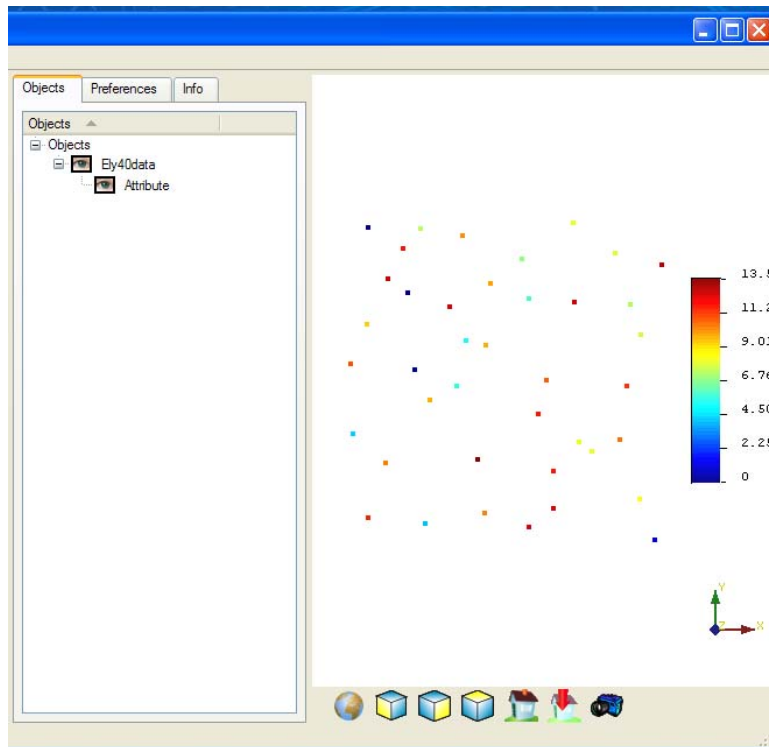
Loading is complete.
To post data, operate on <Object> and <Preferences>.

B. CREATE GRID FOR STORING RESULTS

- Click on <Objects>;
- click on <New Cartesian Grid>;
- fill in options;
- click on <Create Grid>.
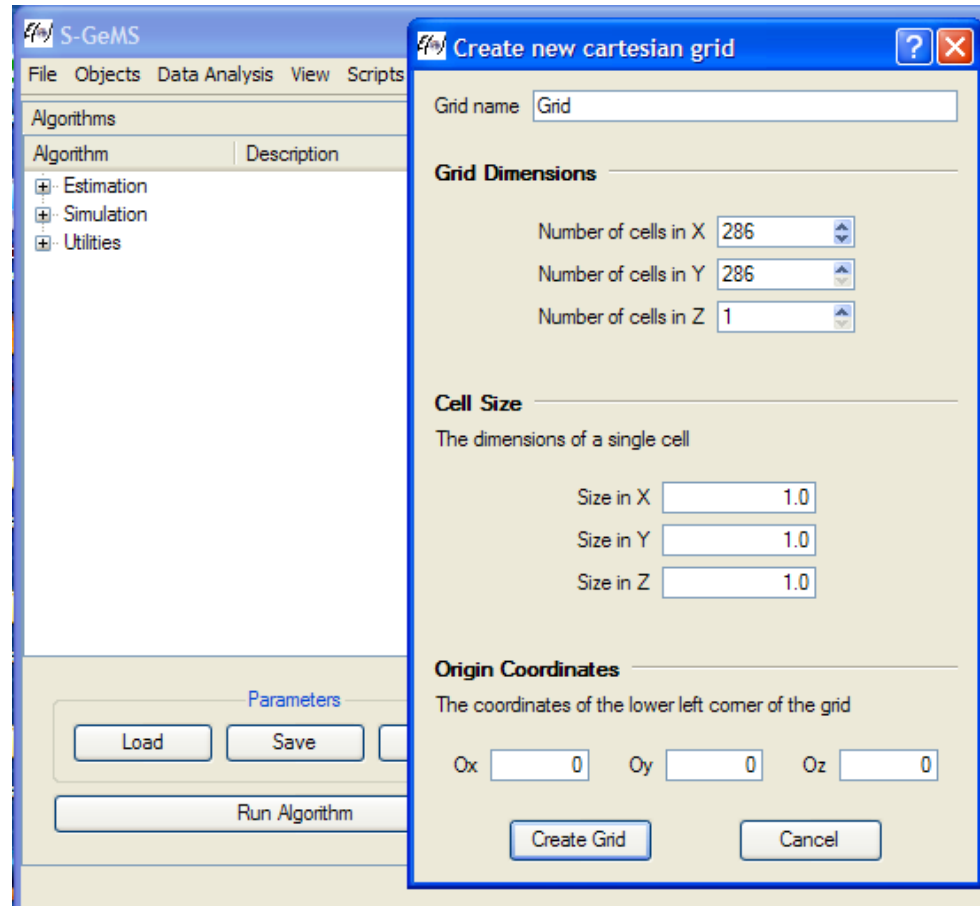
C. SET PARAMETERS

- Click on <Estimation> and then on <kriging>.

- Complete the form. To be able to do that, it is required to previously have modeled the semivariogram.

- Click on <Run Algorithm> when ready.

You may want to click on <Save> to store all parameters as *.par file for further reference.

D. RESULTS

z^*_{OK} σ^2_{OK}
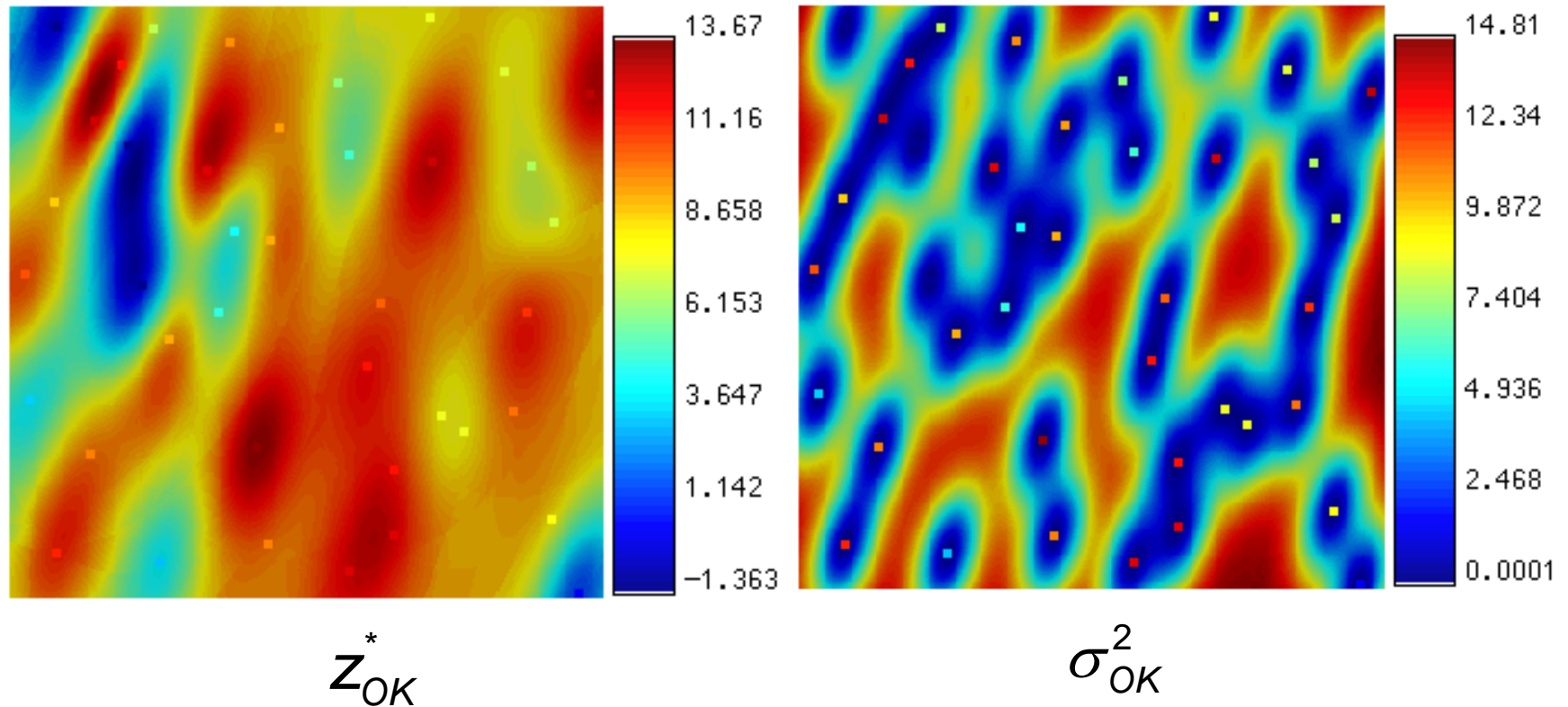
Click on camera icon ( ) at lower right corner of screen
to save maps as electronic files.

10. UNIVERSAL KRIGING

INTRODUCTORY REMARKS

There are several instances of attributes—water depth near the shore, temperature in the upper part of the earth's crust—that have a clear, systematic variation. Hence, models that presume constancy of the mean are inadequate for the characterization of such attributes.

Universal kriging is a full generalization of simple kriging that takes to the limit the ordinary kriging improvement. Universal kriging removes both the requirement that the attribute must have a constant mean and that the user must know such constant mean.

As the model complexity goes up, unfortunately so does uncertainty. Thus universal kriging should not be used indiscriminately.

Universal kriging is also known as kriging with trend.

ESTIMATOR

The universal kriging estimator has the same form as that for ordinary kriging:

$$Z_{UK}^{*}(\mathbf{s}_0) = \begin{cases} \sum_{i=1}^{k} \lambda_i Z(\mathbf{s}_i) \\ \sum_{i=1}^{k} \lambda_i = 1, \end{cases}$$

where

$\mathbf{s}$ denotes spatial location,

λ_i are real weights,

k is the number of observations to consider, and

$Z(\mathbf{s}_i)$ is a variate at site $\mathbf{s}_i$.

ASSUMPTIONS

1. The residuals $Y(\mathbf{s})$ are a partial realization of a random function $Y(\mathbf{s}) = Z(\mathbf{s}) - m_Z(\mathbf{s})$, where $m_Z(\mathbf{s})$ is the trend or drift of the random function $Z(\mathbf{s})$.

2. As in ordinary kriging, one can select second order stationarity or intrinsic hypothesis, depending on whether the derivations will be done on terms of covariances or semivariograms of the residuals.

3. The mean is now neither known nor constant.

4. The trend in the mean is considered a deterministic component amenable to analytical modeling. The most common practice, yet not the only alternative, is to employ polynomials of the geographical coordinates. The coefficients become additional unknowns in the trend expression.

TREND (DRIFT)
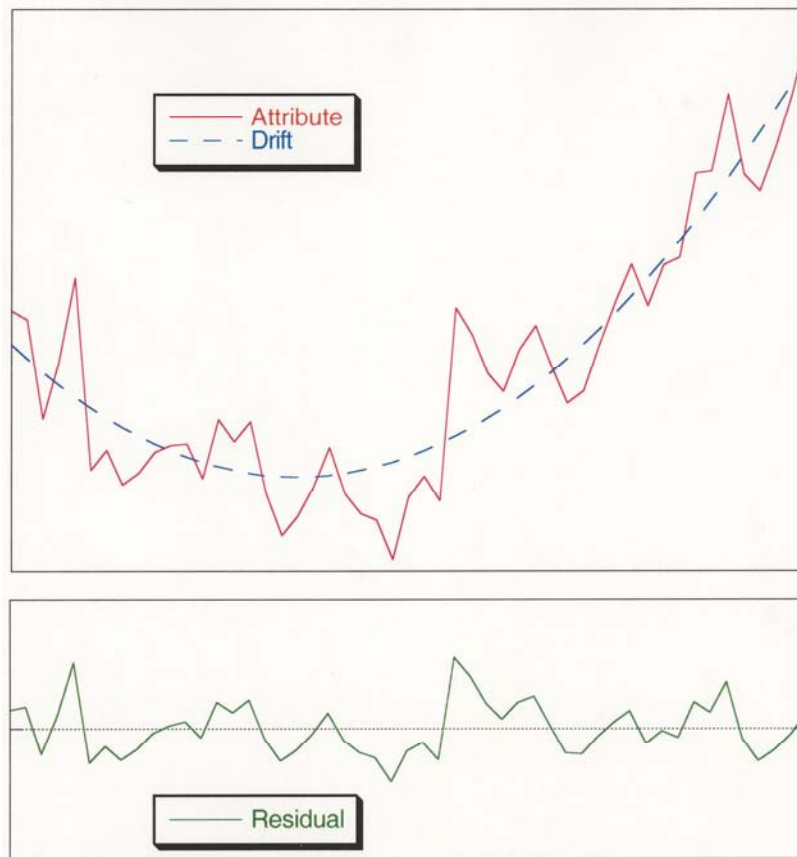
$$m_z(\mathbf{s}) = a_0 + \sum_{i=1}^{n} a_i f_i$$

NORMAL EQUATIONS

$$\sum_{i=1}^{k} \lambda_i \operatorname{Cov}_Y(\mathbf{s}_i, \mathbf{s}_1) + \mu_0 + \sum_{l=1}^{n} \mu_l f_l(\mathbf{s}_1) = \operatorname{Cov}_Y(\mathbf{s}_0, \mathbf{s}_1)$$

$$\sum_{i=1}^{k} \lambda_i \operatorname{Cov}_Y(\mathbf{s}_i, \mathbf{s}_2) + \mu_0 + \sum_{l=1}^{n} \mu_l f_l(\mathbf{s}_2) = \operatorname{Cov}_Y(\mathbf{s}_0, \mathbf{s}_2)$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum_{i=1}^{k} \lambda_i \operatorname{Cov}_Y(\mathbf{s}_i, \mathbf{s}_k) + \mu_0 + \sum_{l=1}^{n} \mu_l f_l(\mathbf{s}_k) = \operatorname{Cov}_Y(\mathbf{s}_0, \mathbf{s}_k)$$

$$\sum_{i=1}^{k} \lambda_i = 1$$

$$\sum_{i=1}^{k} \lambda_i f_1(\mathbf{s}_i) = f_1(\mathbf{s}_0)$$

$$\sum_{i=1}^{k} \lambda_i f_2(\mathbf{s}_i) = f_2(\mathbf{s}_0)$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum_{i=1}^{k} \lambda_i f_k(\mathbf{s}_i) = f_n(\mathbf{s}_0)$$

Note that the covariance is that of the residuals.

ESTIMATION ERROR VARIANCE

$$\sigma_{UK}^2(\mathbf{s}_0) = \text{Cov}_Y(0) - \sum_{i=1}^{k} \lambda_i \, \text{Cov}_Y(\mathbf{s}_i, \mathbf{s}_0) - \mu_0 - \sum_{j=1}^{n} \mu_j f_j(\mathbf{s}_0)$$

SOME MATRIX DEFINITIONS

$$\mathbf{K} = \begin{bmatrix} Cov_Y(\mathbf{s}_1,\mathbf{s}_1) & \cdots & Cov_Y(\mathbf{s}_k,\mathbf{s}_1) & 1 & f_1(\mathbf{s}_1) & \cdots & f_n(\mathbf{s}_1) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Cov_Y(\mathbf{s}_1,\mathbf{s}_k) & \cdots & Cov_Y(\mathbf{s}_k,\mathbf{s}_k) & 1 & f_1(\mathbf{s}_k) & \cdots & f_n(\mathbf{s}_k) \\ 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ f_1(\mathbf{s}_1) & \cdots & f_1(\mathbf{s}_k) & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ f_n(\mathbf{s}_1) & \cdots & f_n(\mathbf{s}_k) & 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$\mathbf{k} = \begin{bmatrix} Cov_Y(\mathbf{s}_0,\mathbf{s}_1) & \cdots & Cov_Y(\mathbf{s}_0,\mathbf{s}_k) & 1 & f_1(\mathbf{s}_0) & \cdots & f_n(\mathbf{s}_0) \end{bmatrix}^T$$

$$\mathbf{x} = \begin{bmatrix} \lambda_1 & \cdots & \lambda_k & -\mu_0 & -\mu_1 & \cdots & -\mu_n \end{bmatrix}^T$$

$$\mathbf{Z} = \begin{bmatrix} Z(\mathbf{s}_1) & \cdots & Z(\mathbf{s}_k) & 0 & 0 & \cdots & 0 \end{bmatrix}^T$$

199

MATRIX FORMULATIONS

Normal equations

$$\mathbf{Kx} = \mathbf{k}$$

Estimator

$$Z_{UK}^*(\mathbf{s}_0) = \mathbf{Z}^T\mathbf{x}$$

Estimation variance

$$\sigma_{UK}^2 = Cov_Y(0) - \mathbf{k}^T\mathbf{x}$$

NUMERICAL EXERCISE

Elevation of water table in northwestern Kansas is a clear example of an attribute with a linear trend. Use the sample to:

A. Apply the semivariogram estimator on page 114 along different directions.
B. Obtain the semivariogram for the residuals by fitting a model to the experimental semivariogram along the trend-free direction—in this case, the only experimental semivariogram that is not an artifact.
C. Use the closest observations to estimate the weights and Lagrange multipliers at location (60, 193).
D. Find the estimate and its kriging variance.
E. Produce a map of water table elevations using SGeMS.

A. EXPERIMENTAL SEMIVARIOGRAM FOR THE ATTRIBUTE
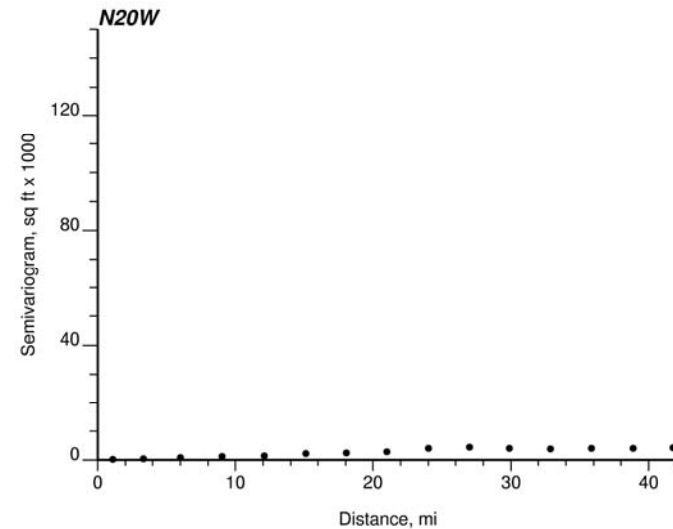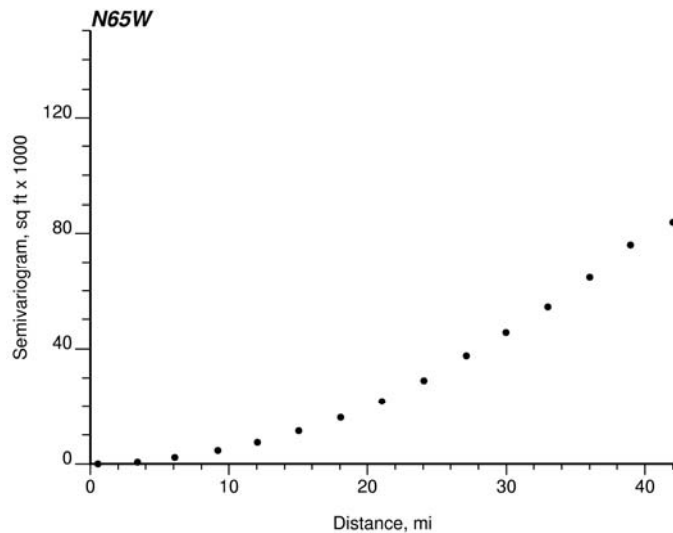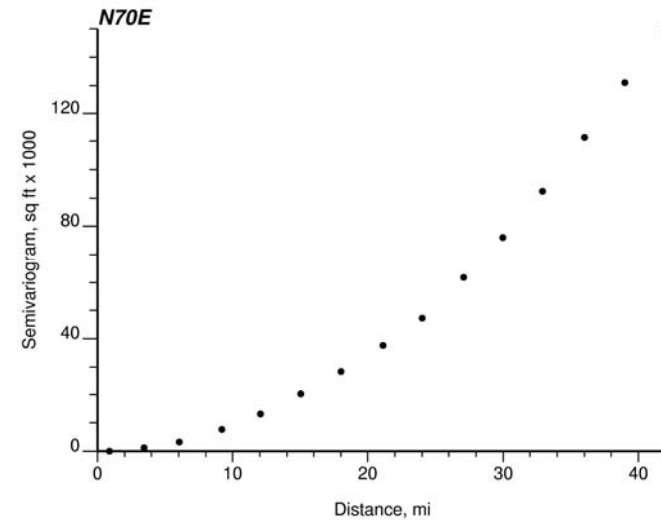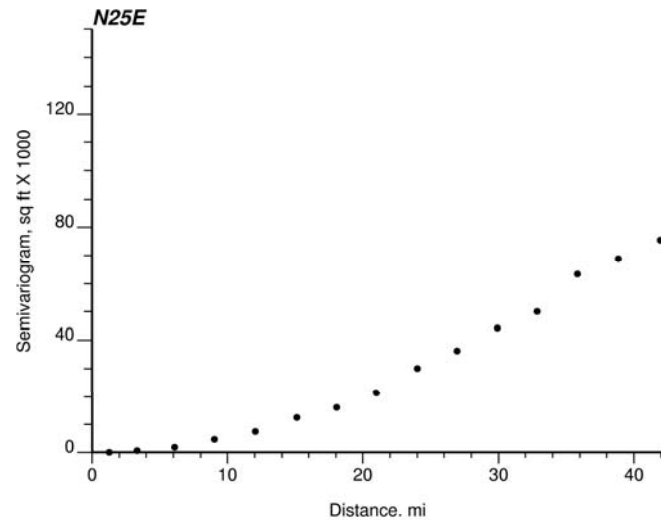
B. SEMIVARIOGRAM MODEL ALONG TREND-FREE DIRECTION
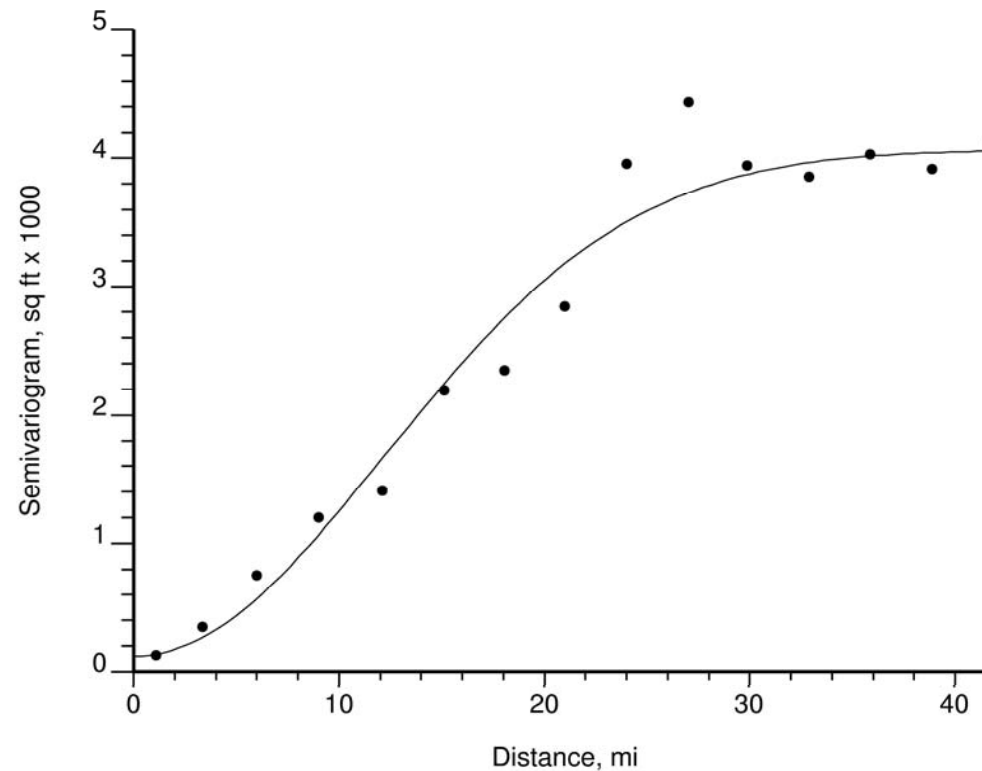
$$\gamma(\mathbf{h}) = 119 + \text{Gaussian}(\text{sill} = 3948, \text{range} = 29.8)$$

C. WEIGHTS
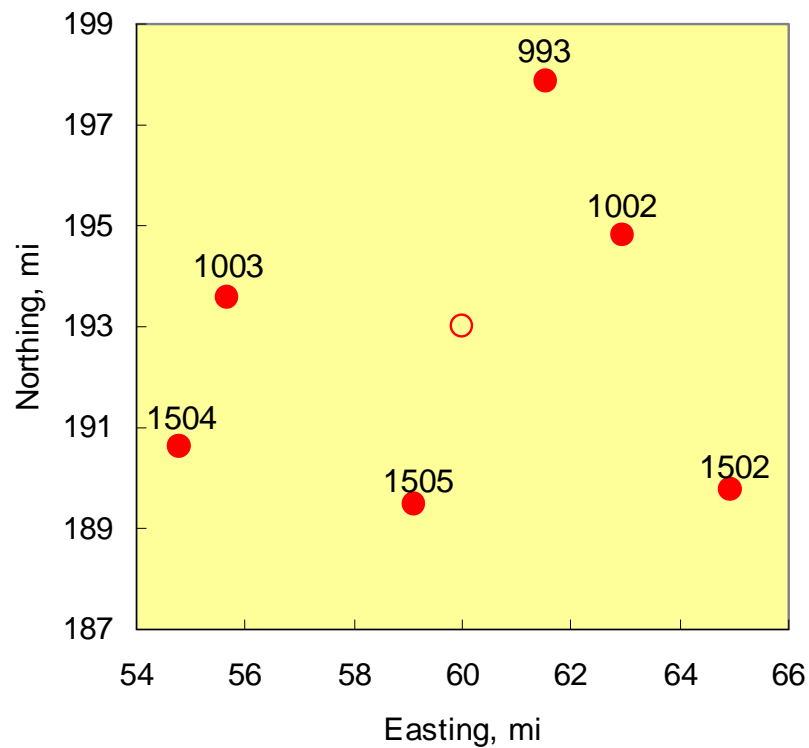
Data

Well	Easting mi	Northing mi	Elevation ft
993	61.56	197.85	3065.0
1002	62.94	194.81	3099.4
1003	55.68	193.56	3200.0
1502	64.96	189.77	3114.9
1504	54.80	190.60	3217.1
1505	59.12	189.47	3189.7
	60.0	193.0	?

DISTANCES

Observation–observation:

$$\begin{bmatrix} 0 & & & & & \\ 3.34 & 0 & & & & \\ 7.27 & 7.37 & 0 & & & \\ 8.77 & 5.44 & 10.03 & 0 & & \\ 9.91 & 9.17 & 3.09 & 10.20 & 0 & \\ 8.72 & 6.57 & 5.34 & 5.85 & 4.47 & 0 \end{bmatrix}$$

Estimate–observation:

$$\begin{bmatrix} 5.09 & 3.45 & 4.36 & 5.92 & 5.73 & 3.64 \end{bmatrix}^T$$

COVARIANCES AND UNKNOWNS

$$
\mathbf{K} = \begin{bmatrix}
3710.5 \\
3489.4 & 3710.5 \\
2970.1 & 2954.3 & 3710.5 \\
2707.1 & 3250.3 & 2471.3 & 3710.5 \\
2494.0 & 2633.9 & 3511.0 & 2438.4 & 3710.5 \\
2716.0 & 3084.4 & 3262.7 & 3192.0 & 3372.7 & 3710.5 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 \\
61.56 & 62.94 & 55.68 & 64.96 & 54.80 & 59.12 & 0 & 0 \\
187.85 & 194.81 & 193.56 & 189.77 & 190.60 & 189.47 & 0 & 0 & 0
\end{bmatrix}
$$

$$
\mathbf{k} = \begin{bmatrix} 3296.1 & 3478.8 & 3385.5 & 3181.3 & 3209.4 & 3461.4 & 1 & 60.0 & 193.0 \end{bmatrix}^T
$$

$$
\mathbf{x} = \begin{bmatrix} 0.073 & 0.348 & 0.274 & 0.012 & -0.056 & 0.350 & 33.95 & 0.27 & 0.53 \end{bmatrix}^T
$$

D. ESTIMATE AND VARIANCE

$$z^*_{UK}(60,193) = \begin{bmatrix} 3065.0 \\ 3099.4 \\ 3200.0 \\ 3114.9 \\ 3217.1 \\ 3189.7 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} 0.073 \\ 0.348 \\ 0.274 \\ 0.012 \\ -0.056 \\ 0.350 \\ 33.95 \\ 0.27 \\ 0.53 \end{bmatrix} \qquad \sigma^2_{UK}(60,193) = \begin{bmatrix} 3296.1 \\ 3478.8 \\ 3385.5 \\ 3181.3 \\ 3209.4 \\ 3461.4 \\ 1 \\ 60.0 \\ 193.00 \end{bmatrix}^T \begin{bmatrix} 0.073 \\ 0.348 \\ 0.274 \\ 0.012 \\ -0.056 \\ 0.350 \\ 33.95 \\ 0.27 \\ 0.53 \end{bmatrix}$$

$$z^*_{UK}(60,193) = 3149.4 \qquad\qquad \sigma^2_{UK}(60,193) = 111.71$$

WEIGHTS, ESTIMATES, AND VARIANCES
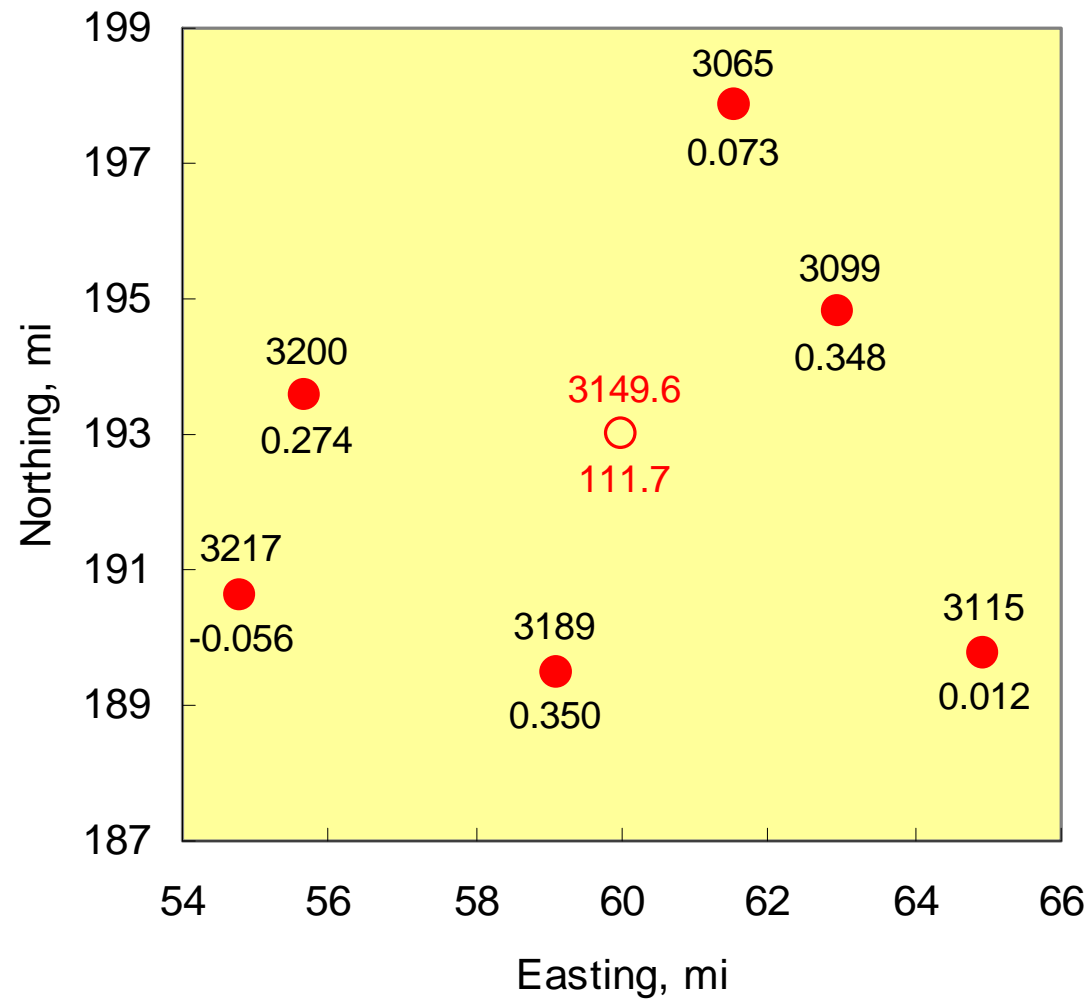
E. WATER TABLE ELEVATION WITH SGeMS

- Click on <Estimation and <kriging>;
- Fill in form and modify default values. Trend model is:

$$m(\mathbf{s}) = a + b \cdot X + c \cdot Y$$

where:

$m(\mathbf{s})$: trend
a, b, c: coefficients
X: easting
Y: northing

- Click on <Run Algorithm>.

MAPS
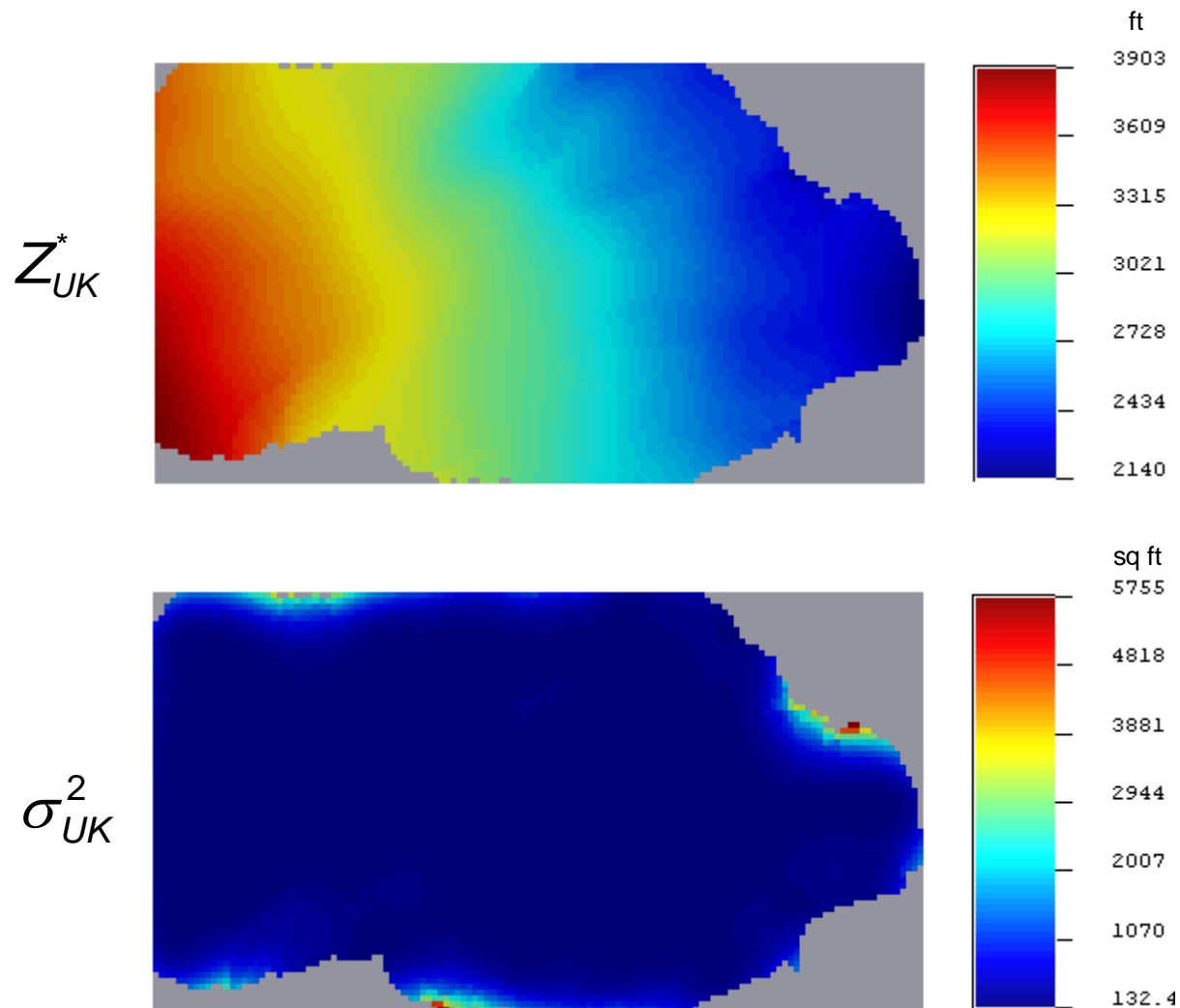
11. BLOCK KRIGING

MOTIVATION

- Block kriging is the generic name given to any form of kriging in which one is interested in estimates of the linear average of an attribute inside supports that are intermediate in size between the support of the observations and the sampling domain.

- This formulation was originally the most widely used form of kriging in mining applications.

- Any form of point kriging is a limiting form of an equivalent formulation of block kriging.

- Ordinary block kriging under second order stationarity is the most common form of block kriging.

LINEAR AVERAGE

Let $Z(\mathbf{s})$ be a point support random function in D and let $S(\mathbf{s}_0)$ be a support of finite size S centered around $\mathbf{s}_0$. The true block average $Z_S(\mathbf{s}_0)$ at location $\mathbf{s}_0$ is:

$$Z_S(\mathbf{s}_0) = \frac{1}{S} \int_{S(\mathbf{s}_0)} Z_S(\mathbf{s})\, d\mathbf{s}.$$

ESTIMATOR

The purpose of block kriging is to produce direct block averages from point or quasi-point measurements, not by averaging point estimates.

Let $Z(\mathbf{s}_i)$ be k variates of a random function and let $S(\mathbf{s}_0) \subset D$ be the support for the estimator $Z_S^*(\mathbf{s}_0)$ at site $\mathbf{s}_0$. Then the ordinary block kriging estimator is:

$$\left. \begin{array}{c} Z_{OBK}^*(\mathbf{s}_0) = \sum_{i=1}^{k} \lambda_i Z(s_i) \\ \sum_{i=1}^{k} \lambda_i = 1 \end{array} \right\}$$

POINT TO BLOCK COVARIANCE

The main peculiarity of block kriging is the use of supports of two different sizes for the observations and the estimator, and hence covariances between two different volume sizes.

Let $\text{Cov}(\mathbf{s}_0, \mathbf{s})$ be covariances between the variate $Z(\mathbf{s}_0)$ and all the variates $Z(\mathbf{s})$ inside the block S. Then the covariance $\text{Cv}(\mathbf{s}_0, S)$ between the site and the block is:

$$\text{Cv}(\mathbf{s}_0, S) = \frac{1}{S} \int_S \text{Cov}(\mathbf{s}_0, \mathbf{s}) \, d\mathbf{s}.$$

BLOCK TO BLOCK COVARIANCE

Let $\mathrm{Cov}(\mathbf{s},\mathbf{u})$ be covariances between a variate $Z(\mathbf{s})$ in block S_i and variates $Z(\mathbf{u})$ inside the block S_j. Then the block covariance $\mathrm{CV}(S_i, S_j)$ between two blocks is:

$$\mathrm{CV}(S_i, S_j) = \frac{1}{S_i S_j} \int\limits_{S_i} \int\limits_{S_j} \mathrm{Cov}(\mathbf{s},\mathbf{u})\, d\mathbf{s}\, d\mathbf{u}.$$

ASSUMPTIONS

The minimum set of assumptions required to formulate ordinary block kriging are the same as those for ordinary kriging.

1. The sample is a partial realization of a random function, $Z(\mathbf{s})$, where $\mathbf{s}$ denotes spatial location.

2. $Z(\mathbf{s})$ is second order stationary, which implies:

$$E[Z(\mathbf{s})] = m$$

$$E[(Z(\mathbf{s}) - m)(Z(\mathbf{s} + \mathbf{h}) - m)] = \text{Cov}(\mathbf{h}),$$

where $E[\cdot]$ denotes expected value, m is a scalar, unknown constant, and $\mathbf{h}$ is a vectorial distance.

3. The mean is constant, but unknown.

NORMAL EQUATIONS

In an entirely similar way to the derivation of the normal equations for point ordinary kriging, one can obtain the following system of equations that provides the ordinary block kriging weights λ_i:

$$
\left.
\begin{array}{l}
\displaystyle\sum_{i=1}^{k} \lambda_i \, \text{Cov}(\mathbf{s}_i, \mathbf{s}_1) + \mu = \text{Cv}(\mathbf{s}_1, S) \\[2.5em]
\displaystyle\sum_{i=1}^{k} \lambda_i \, \text{Cov}(\mathbf{s}_i, \mathbf{s}_2) + \mu = \text{Cv}(\mathbf{s}_2, S) \\[2em]
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\[1em]
\displaystyle\sum_{i=1}^{k} \lambda_i \, \text{Cov}(\mathbf{s}_i, \mathbf{s}_k) + \mu = \text{Cv}(\mathbf{s}_k, S) \\[2.5em]
\displaystyle\sum_{i=1}^{k} \lambda_i = 1
\end{array}
\right\}
$$

ESTIMATION VARIANCE

The minimum mean square error for ordinary block kriging, σ^2_{OBK} , is:

$$\sigma^2_{OBK}(\mathbf{s}_0) = CV(S,S) - \sum_{i=1}^{k} Cv\, \lambda_i(\mathbf{s}_i, S) - \mu$$

The form of the expression remains the same as the one for point ordinary kriging, but all point covariances have been replaced by block-to-block or point-block covariances.

POINT COVARIANCE MODELING

As in other forms of kriging, performance of block kriging presumes knowledge of some covariance terms, which in this case, however, are of different types.

The solution to this requirement is not different from what we have seen before: model the covariances and ignore the effects that the use of models instead of the true covariances have in the normal equations. The common practice is to first model the semivariogram and then use

$$\mathrm{Cov}(\mathbf{h}) = \mathrm{Cov}(0) - \gamma(\mathbf{h})$$

to obtain the point-point covariance model. The block-block and point-block covariances are derived from the point-point covariance.
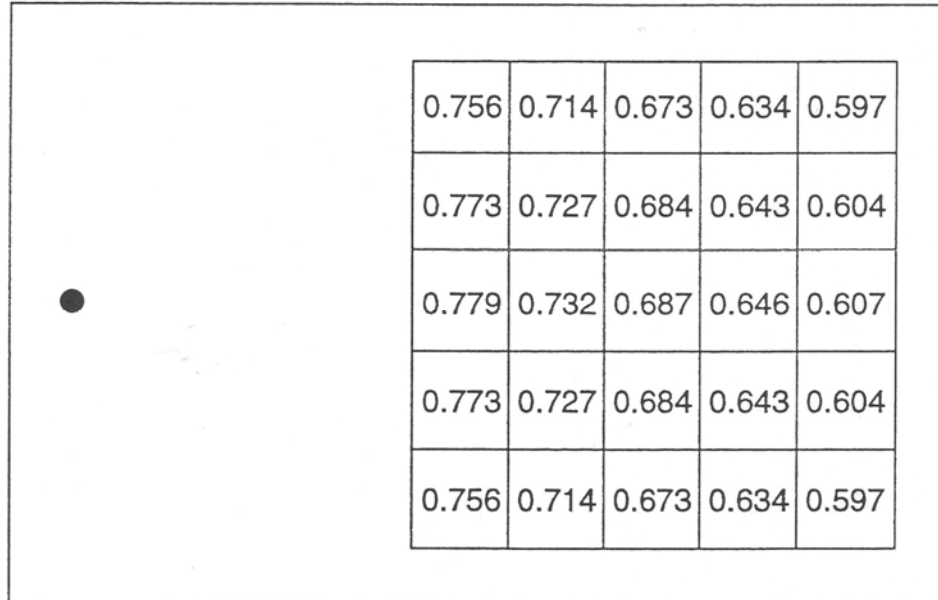
NON-POINT COVARIANCE MODELING

In early days of geostatistics there was great effort devoted to solve exactly the integrals for point-block and block-block covariances for multiples shapes of blocks.

Today the simplest and prevailing practice is to calculate the integrals by numerical approximation. The block is tessellated into smaller units that are treated as points and then the block covariance is calculated as the average between all possible point-point covariances. The following table gives the recommended minimum number of subdivisions.

Dimension	Subdivisions
1	10
2	6 by 6
3	4 by 4 by 4

EXAMPLE OF NUMERICAL CALCULATION

0.756	0.714	0.673	0.634	0.597
0.773	0.727	0.684	0.643	0.604
0.779	0.732	0.687	0.646	0.607
0.773	0.727	0.684	0.643	0.604
0.756	0.714	0.673	0.634	0.597

The values show the point covariance between the point on the left and the center of the 0.05 mile squares of the subdivision. Averaging the 25 numbers gives a value of 0.682 for Cv(0.3).

AVERAGE POROSITY
WEST LYONS FIELD, KANSAS

Ordinary kriging

Block kriging
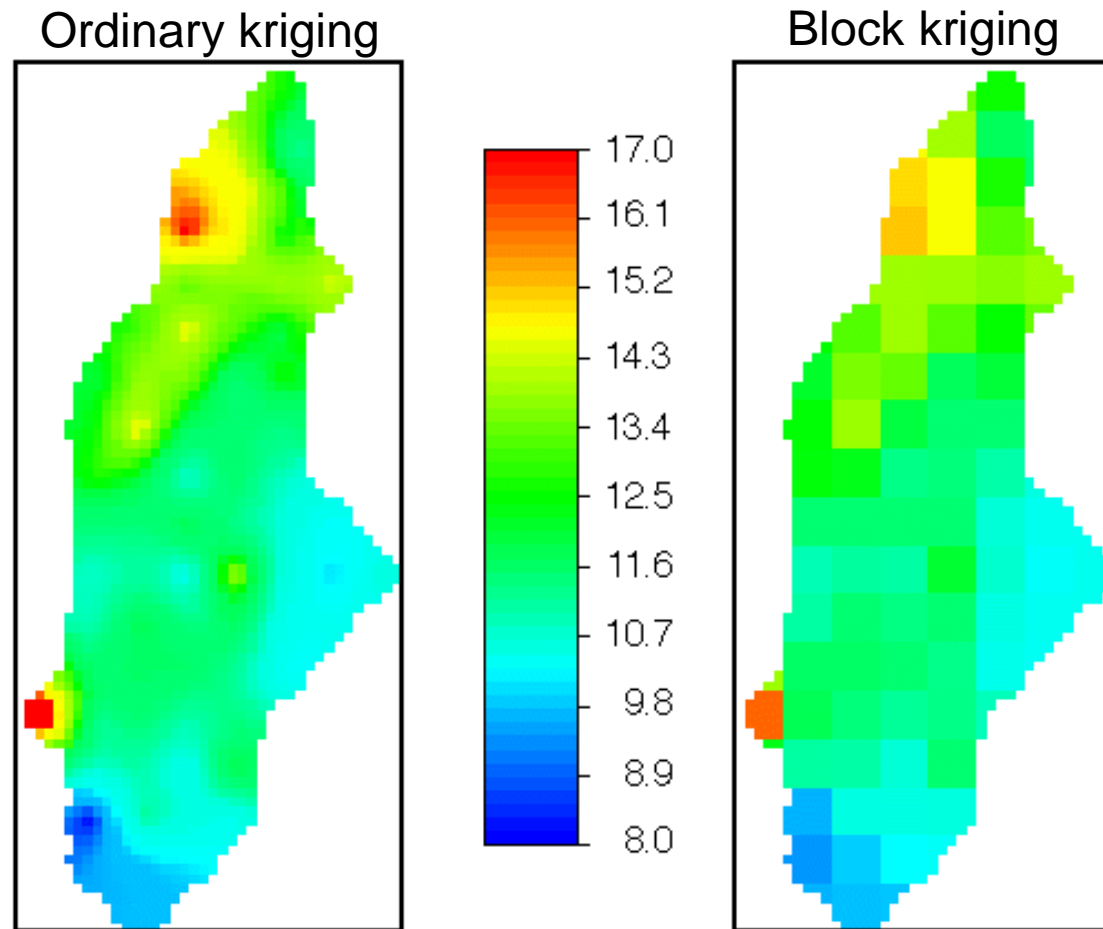
STANDARD ERROR OF
NORMAL SCORES OF POROSITY

Ordinary kriging

Block kriging
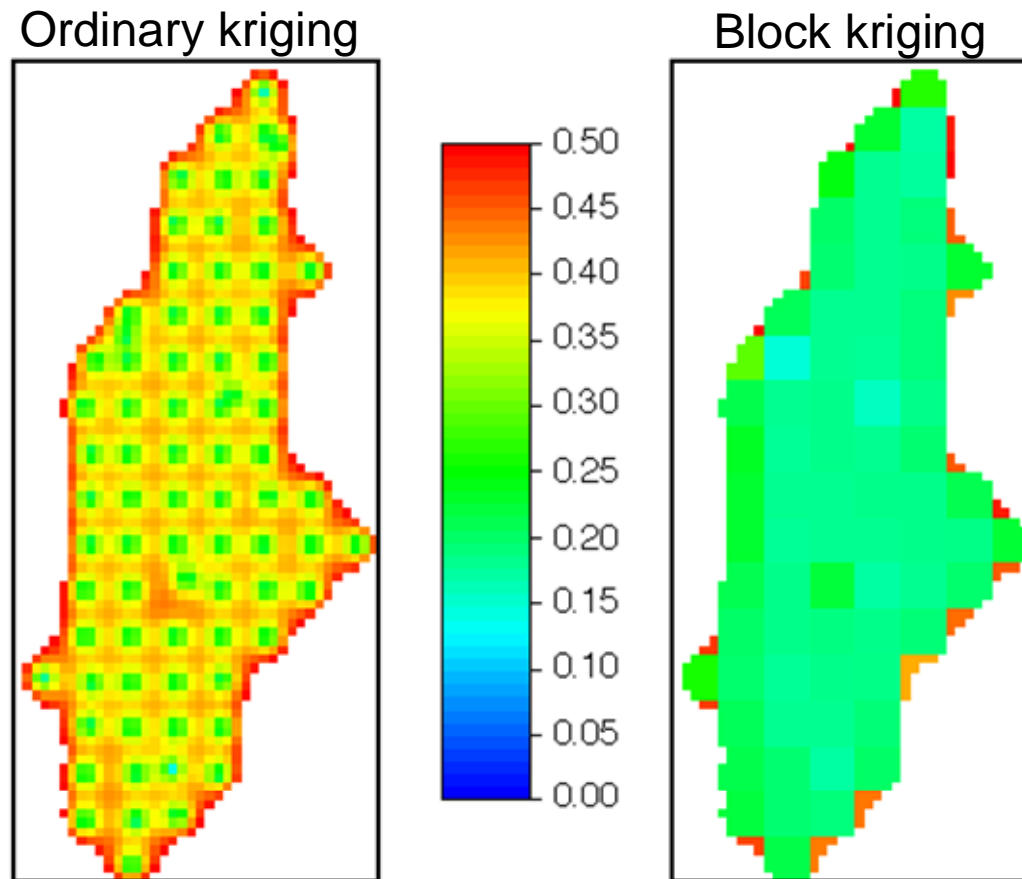
12. COKRIGING

INTRODUCTION

- Simple, ordinary, and universal kriging are not multivariate models in the usual statistical sense of the term. Despite the fact that they employ random function models comprising infinite numbers of random variables, they are employed in the modeling of one attribute at a time.

- The main purpose of cokriging is to work as a true multivariate estimation method able to deal simultaneously with two or more attributes defined over the same domain, which is called a coregionalization.

- Not all attributes must be measured at all locations, but a bare minimum of collocated measurements per pair of attributes is necessary for structural analysis.

VECTORIAL NOTATION

Like in spatial univariate geostatistics, geographical location is the vectorial term:

$$\mathbf{s}_i = \left[\left(easting_i,\ northing_i,\ elevation_i\right)\right]^T.$$

Cokriging makes heavy use of vectorial and matrix notation for conciseness in the notation and similarity with the univariate formulations. Let us start with the definition of a vectorial random function.

If $Z_1(\mathbf{s}), Z_2(\mathbf{s}), \ldots, Z_k(\mathbf{s}), \ldots, Z_p(\mathbf{s})$ is a set of p random functions defined over the same domain, then the vectorial random function $\mathbf{Z}(\mathbf{s})$ is the vectorial matrix:

$$\mathbf{Z}(\mathbf{s}) = \left[Z_1(\mathbf{s})\ Z_2(\mathbf{s}) \cdots Z_k(\mathbf{s}) \cdots Z_p(\mathbf{s})\right]^T.$$

COEFFICIENT MATRIX

The coefficient matrix, Λ_i, is the square matrix:

$$\Lambda_i = \begin{bmatrix} \lambda_{11}^i & \lambda_{12}^i & \cdots & \lambda_{1p}^i \\ \lambda_{21}^i & \lambda_{22}^i & \cdots & \lambda_{2p}^i \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{p1}^i & \lambda_{p2}^i & \cdots & \lambda_{pp}^i \end{bmatrix}$$

where λ_{jk}^i are real numbers and p is the number of attributes to model.

ORDINARY COKRIGING ESTIMATOR

The objective of cokriging is to find the minimum mean-square-error weights for the linear estimator, $\mathbf{Z}^*(\mathbf{s}_0)$:

$$\mathbf{Z}^*(\mathbf{s}_0) = \sum_{i=1}^{n} \Lambda_i \mathbf{Z}(\mathbf{s}_i)$$

or

$$\mathbf{Z}^*(\mathbf{s}_0) = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \\ \vdots \\ \Lambda_n \end{bmatrix}^T \begin{bmatrix} \mathbf{Z}(\mathbf{s}_1) \\ \mathbf{Z}(\mathbf{s}_2) \\ \vdots \\ \mathbf{Z}(\mathbf{s}_n) \end{bmatrix}$$

both subject to

$$\sum_{i=1}^{n} \Lambda_i = \mathbf{I}.$$

229

VECTORIAL MOMENTS

For the purpose of listing the assumptions, we need to
define the vector of means, **m**:

$$\mathbf{m} = \begin{bmatrix} m_1 & m_2 & \dots & m_p \end{bmatrix}^T$$

and the matrix covariance

$$\mathbf{Cov}(\mathbf{Z}(\mathbf{s}), \mathbf{Z}(\mathbf{s}+\mathbf{h})) = \begin{bmatrix} \text{Cov}_{11}(\mathbf{Z}_1(\mathbf{s}), \mathbf{Z}_1(\mathbf{s}+\mathbf{h})) & \cdots & \text{Cov}_{1p}(\mathbf{Z}_1(\mathbf{s}), \mathbf{Z}_p(\mathbf{s}+\mathbf{h})) \\ \vdots & \vdots & \vdots \\ \text{Cov}_{p1}(\mathbf{Z}_p(\mathbf{s}), \mathbf{Z}_1(\mathbf{s}+\mathbf{h})) & \cdots & \text{Cov}_{pp}(\mathbf{Z}_p(\mathbf{s}), \mathbf{Z}_p(\mathbf{s}+\mathbf{h})) \end{bmatrix}$$

ORDINARY COKRIGING ASSUMPTIONS

The minimum set of assumptions required to formulate ordinary cokriging under second order stationarity are:

1. The sample is a partial realization of a vectorial random function, $\mathbf{Z}(\mathbf{s})$.

2. $\mathbf{Z}(\mathbf{s})$ is second order stationary, which implies:

$$E\big[\mathbf{Z}(\mathbf{s})\big] = \mathbf{m}$$

$$E\big[(\mathbf{Z}(\mathbf{s}) - \mathbf{m})(\mathbf{Z}(\mathbf{s} + \mathbf{h}) - \mathbf{m})\big] = \mathbf{Cov}(\mathbf{h}).$$

3. Each term in $\mathbf{m}$ is unknown, but constant.

4. None of the variables is a linear combination of the others:

$$Z_k(\mathbf{s}) \neq a_0 + \sum_{j \neq k} a_j Z_j(\mathbf{s}), \ \text{for } k = 1, 2, \ldots, p.$$

NORMAL EQUATIONS

$$\left.\begin{array}{c} \displaystyle\sum_{i=1}^{k} \boldsymbol{\Lambda}'_1 \mathbf{Cov}\left(\mathbf{s}_i, \mathbf{s}_1\right) + \boldsymbol{\mu} = \mathbf{Cov}\left(\mathbf{s}_1, \mathbf{s}_0\right) \\[2ex] \displaystyle\sum_{i=1}^{k} \boldsymbol{\Lambda}'_2 \mathbf{Cov}\left(\mathbf{s}_i, \mathbf{s}_2\right) + \boldsymbol{\mu} = \mathbf{Cov}\left(\mathbf{s}_2, \mathbf{s}_0\right) \\[2ex] \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\[2ex] \displaystyle\sum_{i=1}^{k} \boldsymbol{\Lambda}'_p \mathbf{Cov}\left(\mathbf{s}_i, \mathbf{s}_p\right) + \boldsymbol{\mu} = \mathbf{Cov}\left(\mathbf{s}_p, \mathbf{s}_0\right) \\[2ex] \displaystyle\sum_{i=1}^{n} \boldsymbol{\Lambda}_i = \mathbf{I} \end{array}\right\}$$

where $\boldsymbol{\mu}$ is a square matrix of Lagrange multipliers of the same order p as the number of attributes in the coregionalization.

ESTIMATION VARIANCE

If

$$\mathbf{W} = \begin{bmatrix} \Lambda_1^{'} & \Lambda_2^{'} & \cdots & \Lambda_n^{'} & \boldsymbol{\mu} \end{bmatrix}^T$$

$$\mathbf{k} = \begin{bmatrix} \mathbf{Cov}(\mathbf{s}_1, \mathbf{s}_0) & \cdots & \mathbf{Cov}(\mathbf{s}_n, \mathbf{s}_0) & \mathbf{I} \end{bmatrix}^T,$$

then the estimation variance for ordinary cokriging $\sigma_{OC}^2(\mathbf{s}_0)$ is:

$$\sigma_{OC}^2(\mathbf{s}_0) = \text{diagonal}\left(\mathbf{Cov}(\mathbf{0}) - \mathbf{k}^T \mathbf{W}\right).$$

STRUCTURAL ANALYSIS

The structural analysis required by cokriging is much more demanding than the one for kriging because:

- Besides the p covariances, the user must model additional $p(p-1)/2$ cross covariances.

- $\mathbf{Cov}(\cdot)$ must be positive definite to properly solve the normal equations. Individual selection of permissible models for all covariances and cross covariances is not sufficient to produce a positive definite $\mathbf{Cov}(\cdot)$.

The common way to conduct structural analysis is through the application of the so-called linear coregionalization model.

LINEAR COREGIONALIZATION MODELING

Linear coregionalization modeling is a solution based on the decomposition of all structural functions in terms of a basic set of permissible models, first done in terms of semivariograms and cross semivariograms and then converted to covariances and cross covariances.

$$\Gamma(\mathbf{h}) = \sum_{l=1}^{L} \mathbf{B}_l \gamma_l(\mathbf{h})$$

For all structural functions to be permissible, all coefficient matrices $\mathbf{B}_l$ of order p must be positive semidefinite. If a matrix is positive semidefinite, its determinant, all its principal minor determinants, and all its eigenvalues are nonnegative. A cross semivariogram, $\gamma_{ij}(\mathbf{h})$, is the moment:

$$\gamma_{ij}(\mathbf{h}) = \frac{1}{2} E\left[\left\{Z_i(\mathbf{s}) - Z_i(\mathbf{s}+\mathbf{h})\right\}\left\{Z_j(\mathbf{s}) - Z_j(\mathbf{s}+\mathbf{h})\right\}\right].$$

EXAMPLE OF LINEAR COREGIONALIZATION MODEL

The following would be an example of isotropic linear coregionalization for the case of two attributes considering nugget effect, one exponential model with a range of 200, and one spherical model with a range of 150:

$$\Gamma(\mathbf{h}) = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} + \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix} \left(1 - e^{-\frac{3h}{200}} \right) + \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \left(\frac{3}{2} \frac{h}{150} - \frac{1}{2} \left(\frac{h}{150} \right)^3 \right).$$

This notation implies, for example, that the cross semivariogram between attributes 1 and 2 is:

$$\gamma_{12}(\mathbf{h}) = 1 + 2 \left(1 - e^{-\frac{3h}{200}} \right) + \left(\frac{3}{2} \frac{h}{150} - \frac{1}{2} \left(\frac{h}{150} \right)^3 \right).$$

EFFECTIVE APPLICATION OF COKRIGING

Abundant information on additional variables is not a guarantee of superior results of multivariate cokriging over just kriging for the variable of interest. The following are two prerequisites:

- The secondary variables must be significantly correlated to the primary variable.

- A significant proportion of secondary measurements must be at locations other than those of the primary variable.

A corollary is that when all attributes are measured at the same sites without missing values, cokriging provides similar results to kriging of individual attributes.

POROSITY OF SEA FLOOR, MECKLENBURG BAY, GERMANY
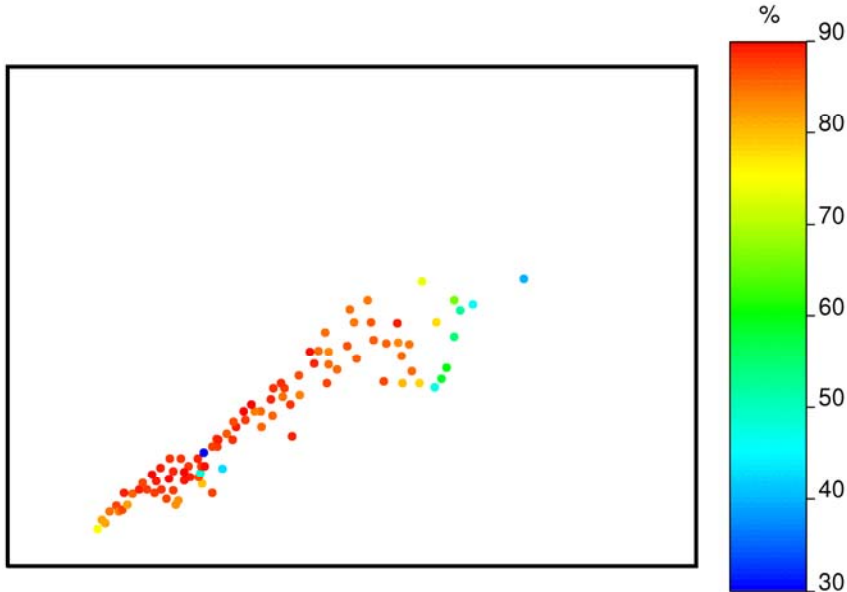
Sample

Kriging estimation
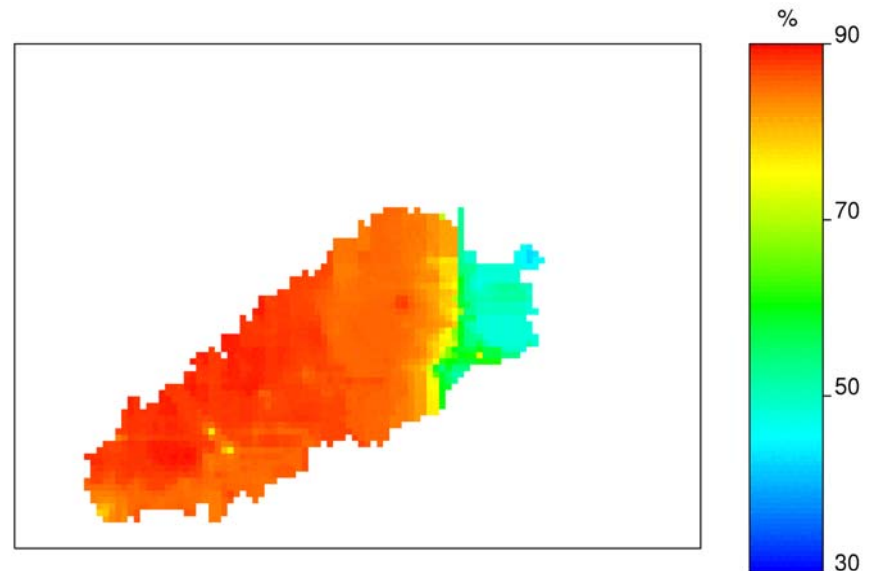
COKRIGING FOR SEA-FLOOR POROSITY MECKLENBURG BAY, GERMANY

Sample for grain size median
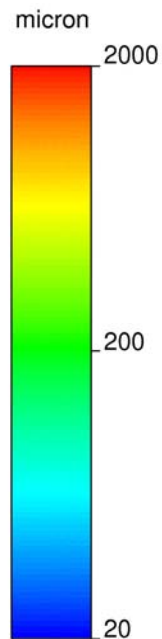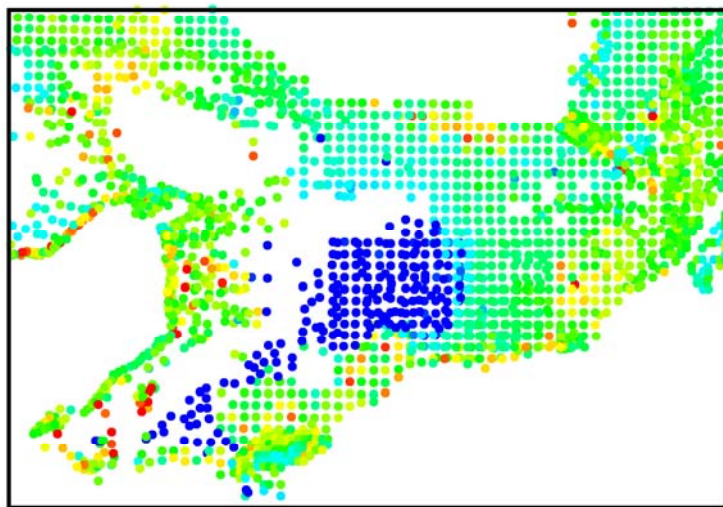
Cokriging porosity map using porosity and grain size median data
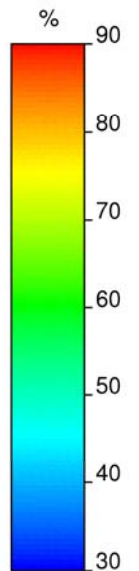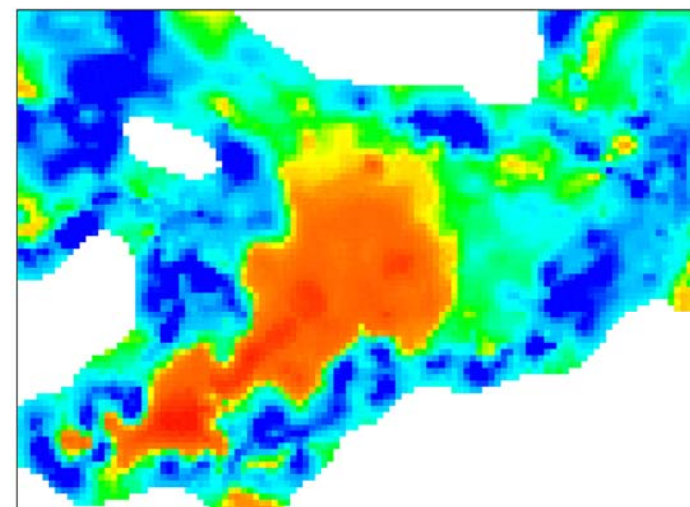
13. CROSSVALIDATION

ACCOUNTABILITY

- Upon selecting any estimator plus its parameters and producing estimations for a given spatial attribute, crossvalidation addresses the curiosity and need to know about the quality of the job done.

- The ultimate approach would be to run an exhaustive sampling and then assess the quality of the estimation by comparing the new data with the collocated estimates produced with the original data. Unquestionably, the best estimator would be the one with the best statistic on the basis of a previously agreed criterion.

- As conclusive as this approach may be, its implementation ordinarily is impossible to prohibitively expensive and defeats the purpose of the estimation in practical situations.

ALTERNATIVE EVALUATION METHOD

Crossvalidation is an ingenious alternative to exhaustive sampling.

- Given a sample of size n, crossvalidation is based on the dropping of each observation with replacement. For each ignored observation, an estimate is computed at the location of the discarded observation by using at most the remaining $n-1$ measurements.

- By pretending that a drawn observation was never taken, one can genuinely attempt to produce an estimate, and by comparing it to the deleted measurement, one can have an instant and cost free error measurement.

- The discarded observation is replaced in the sample before continuing with the analysis.

WORD OF CAUTION

Crossvalidation is a fast and inexpensive way for indirectly testing anything about kriging or the data. However, one must apply crossvalidation with caution.

- Crossvalidation does not indicate whether an observation, estimate, parameter, or assumption is inadequate.

- Complex interdependencies among the errors have precluded finding their distribution, thus hindering a rigorous analysis or testing of the results.

- Nonetheless, crossvalidation remains a useful tool primarily to dispel blunders and assist in drawing honest conclusions by comparison, instead of making arbitrary decisions or assumptions.

   Crossvalidation is not part of SGeMS, but it can be found in GSLIB.

UNFAVORABLE CIRCUMSTANCES

Special circumstances deserving extra caution are:

- In the case of kriging, the estimate is insensitive to scaling of the semivariogram, thus making the estimates insensitive to proportional changes in both sill and nugget.

- Irregular sampling patterns may lead to unrepresentative errors. A typical situation is the sampling in clusters, which results in abnormally low errors at all locations making the clusters. For better results, decluster first, then crossvalidate.

USGS

DIAGNOSTIC STATISTICS

Comparisons or remedial actions based on crossvalidation are more straightforward and conclusive if the collection of errors is reduced to some key statistics such as:

- Some quantiles of the error distribution, such as the 5th percentile, the median, and the 95th percentile.
- The mean and standard deviation for the square of the errors.
- The slope and intercept of the regression line between the measured and estimated values.
- The correlation coefficient between the measured and estimated values.

There is plenty of redundancy in the statistics. In addition, all statistics do not always point in the same direction.

CRITERIA

Guidelines to follow based on the statistics include:

- The ideal estimation method should result in no errors and a correlation coefficient between measurements and estimates of 1.

- Large individual errors may be the effect of blunders or outliers. Data preparation and coding, including geographical location, should be checked carefully.

- A slope different from 45° in the regression line between measurement and estimates denotes conditional bias.

- Ideally, errors should be spatially uncorrelated, resulting in a pure nugget effect semivariogram.

- If the z-scores—the ratio of the error over the standard error—follow a standard normal distribution, one cannot discard the possibility that the errors are multinormally distributed.

PRACTICAL EXAMPLE: UNCF
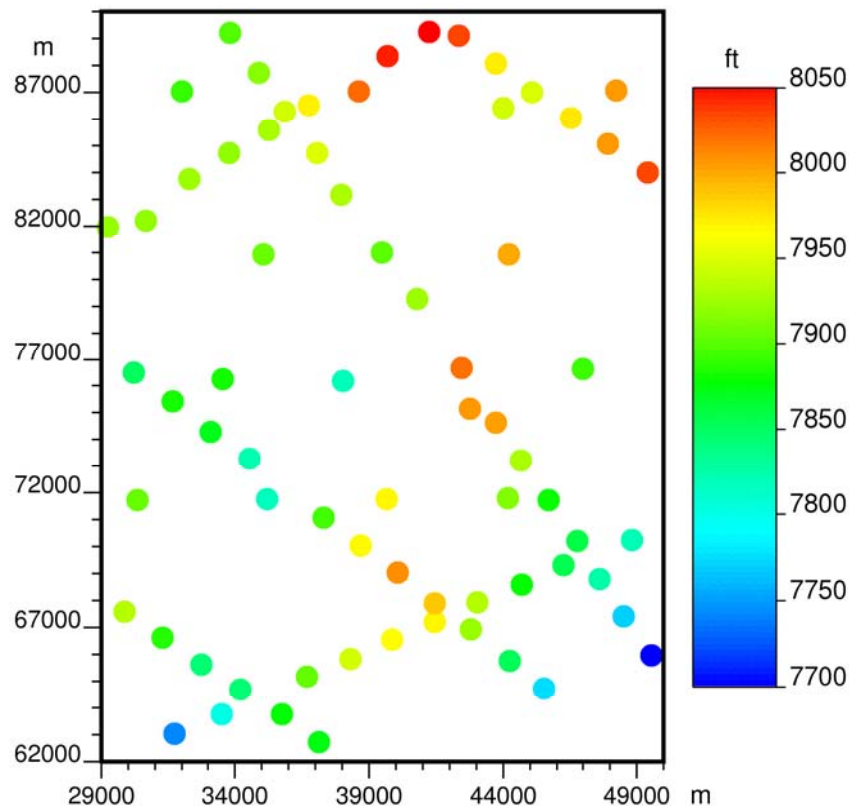
A. The sample size is too small for a directional study of the semivariogram. Employ crossvalidation for investigating anisotropy.

B. Utilize optimal parameters to run ordinary kriging.

C. For an 8-km search radius, employ crossvalidation for finding the optimal number of estimation points.

D. Analyze crossvalidation errors.

A. OPTIMAL NUMBER ESTIMATION POINTS

For an 8-km search radius, the mean square error of ordinary kriging is minimum when the maximum number of estimation points is 11.

Maximum number of observations	Mean square error sq ft
3	589.6
4	632.1
5	658.7
6	657.4
7	593.9
8	545.2
9	539.5
10	451.0
11	361.2
12	376.0
13	385.1
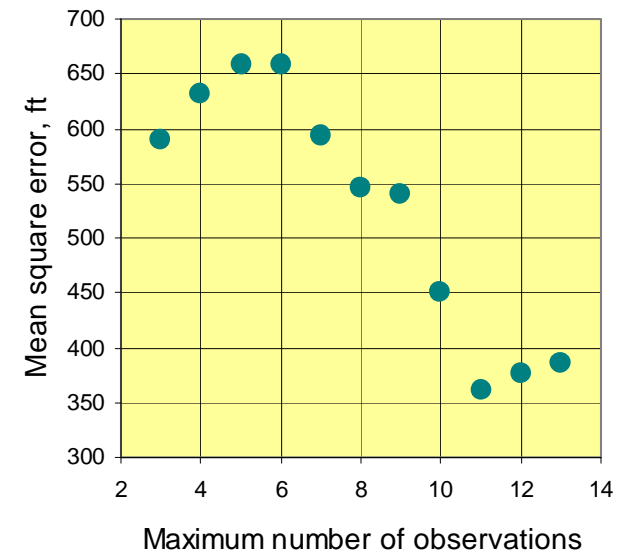
B. INVESTIGATION OF ANISOTROPY

For anisotropic semivariograms, the practice is to vary the range as the length of a cord from the center to the perimeter of an ellipse. Below are the results of an analysis keeping the average of the axes equal to isotropic range.

Isotropic semivariogram
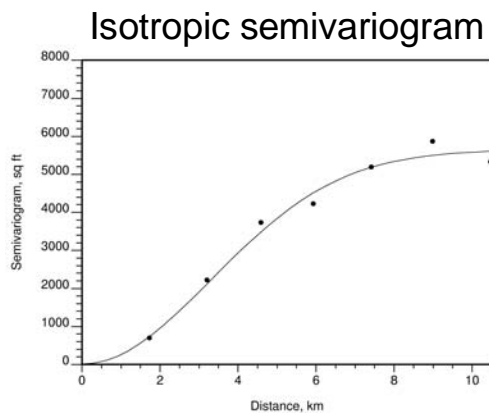
$$\gamma(\mathbf{h}) = 10 + 5.64\left(1 - e^{-\left(\frac{3h}{8.11}\right)^2}\right)$$

Axes size m	Major axis orientation	Mean square error, sq ft
8107 8107	N	361.2
8500 7714	N	302.8
9000 7214	N	257.7
9300 6914	N	247.0
9400 6810	N	246.0
9500 6714	N	246.2
9600 6614	N	247.5
9400 6814	N20E	253.8
9400 6814	N10E	243.6

C. OPTIMAL ORDINARY KRIGING PARAMETERS

- Minimum number of observations: 3
- Maximum number of observations: 11
- Search neighborhood radius: 8,000 m
- Best semivariogram model

Type:	Gaussian
Nugget:	10 sq ft
Sill:	5,640 sq ft
Major range:	9,400 m
Direction:	N10E
Minor range:	6,814 m

ORDINARY KRIGING MAPS
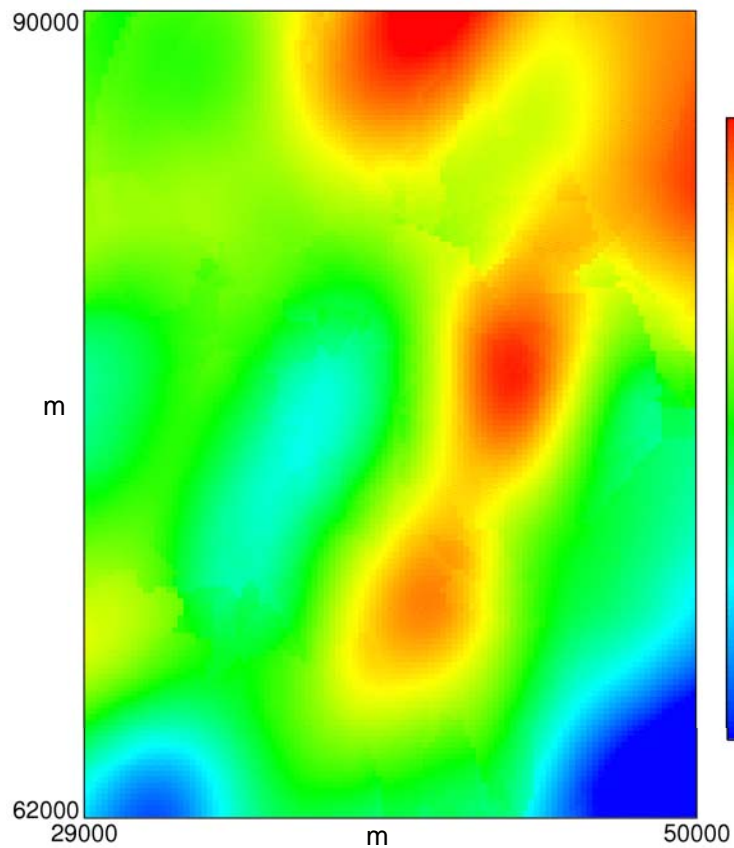
Estimated depth
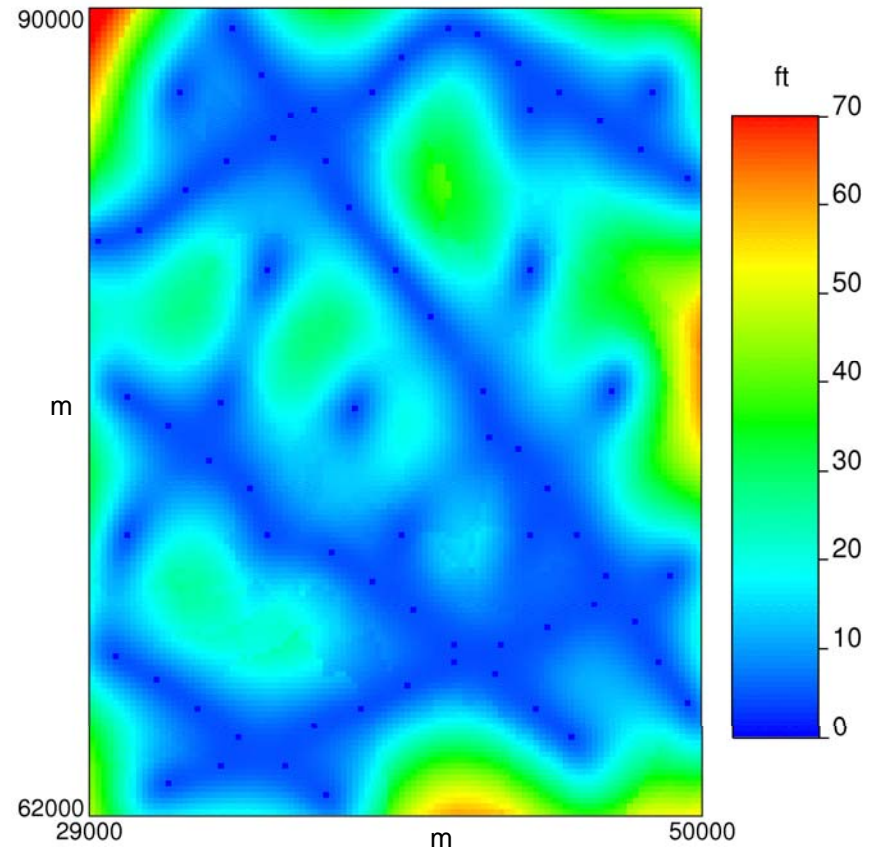
Kriging standard deviation

D. CROSSVALIADATION ERRORS (1)
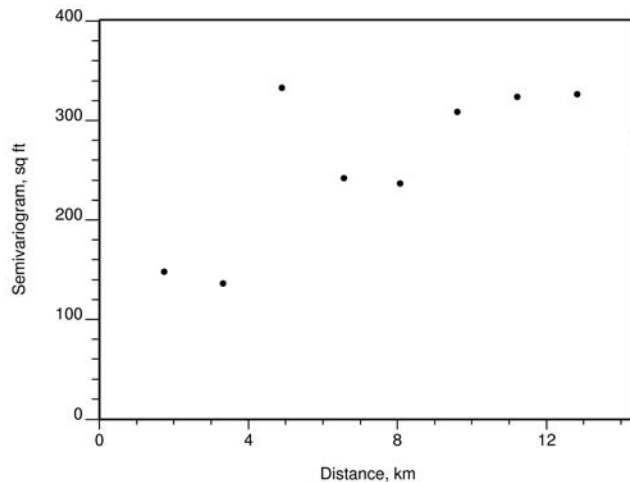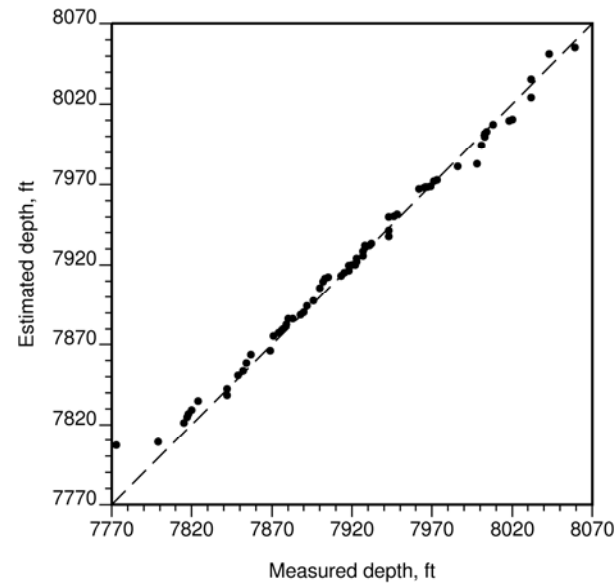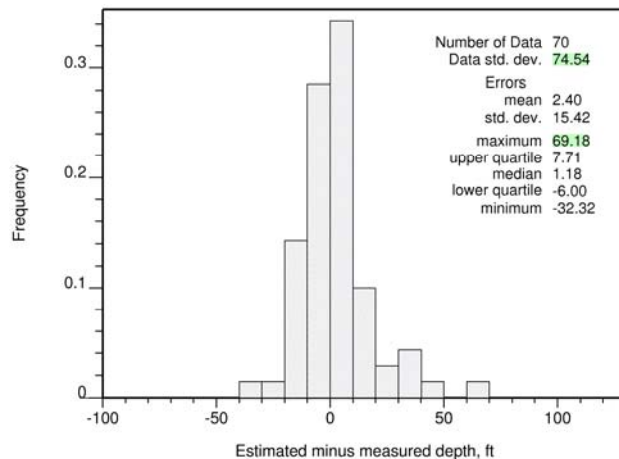
- Errors are slightly positive.
- Distributions for estimated and true values are extraordinarily similar.
- Crossvalidation errors are spatially correlated.

CROSSVALIDATION ERRORS (2)
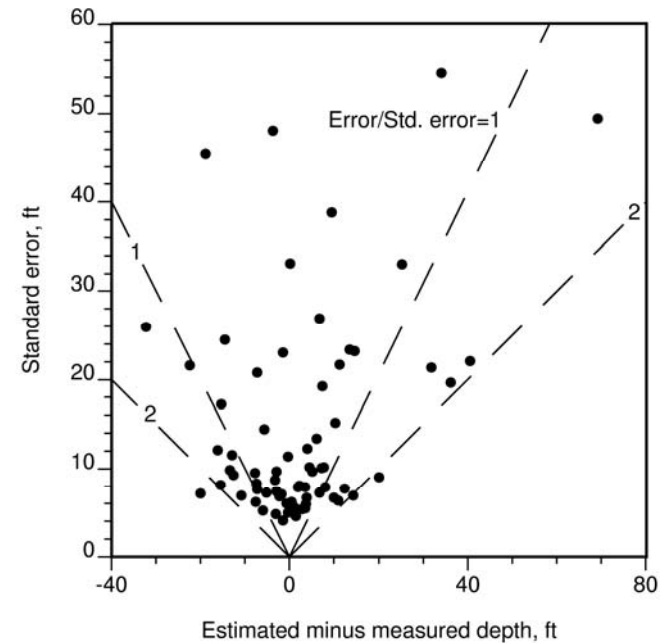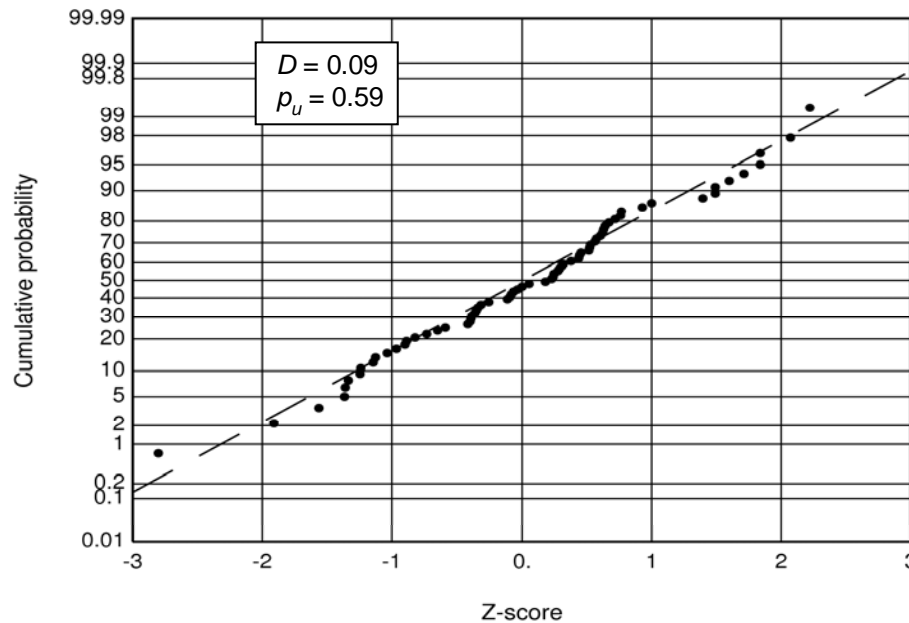
- The Kolmogorov-Smirnov p_u-value is 0.59, which is significantly larger than 0.1. Thus, one cannot reject the normality of errors.
- The larger the standard error, the larger the error dispersion, but without being correlated.

14. CRITICAL REVIEW

SAMPLING

- Minimum mean square error methods—kriging included—produce realistic estimations when the average sampling interval is one order of magnitude smaller than the minor axis of the semivariogram range.

- As the sampling becomes sparser, maps based on kriging show increasing smoothing, which can be appreciated in various ways.

  We explore here the weak sides of kriging as a motivation to go into stochastic simulation.

EVALUATION OF ORDINARY KRIGING THROUGH THE EXHAUSTIVE SAMPLE AT ELY WEST
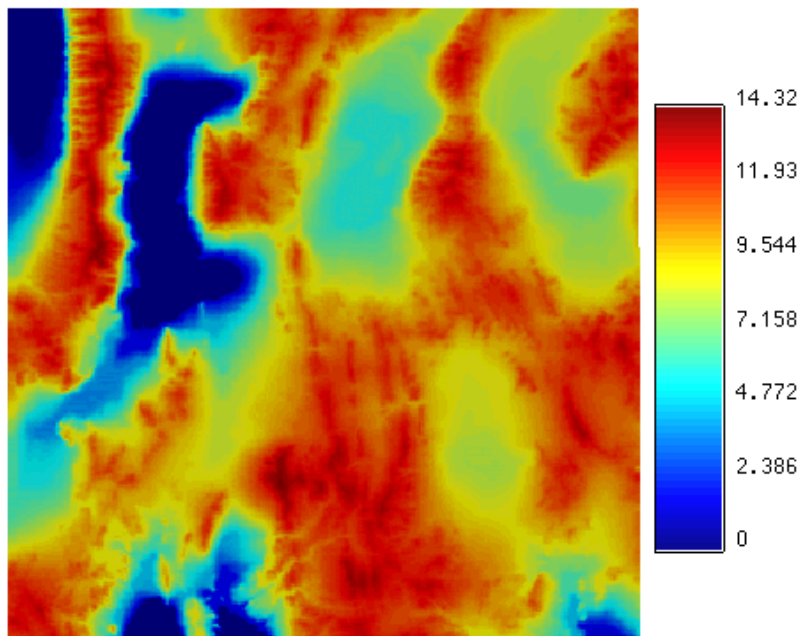
Exhaustive sample
81,796 measurements
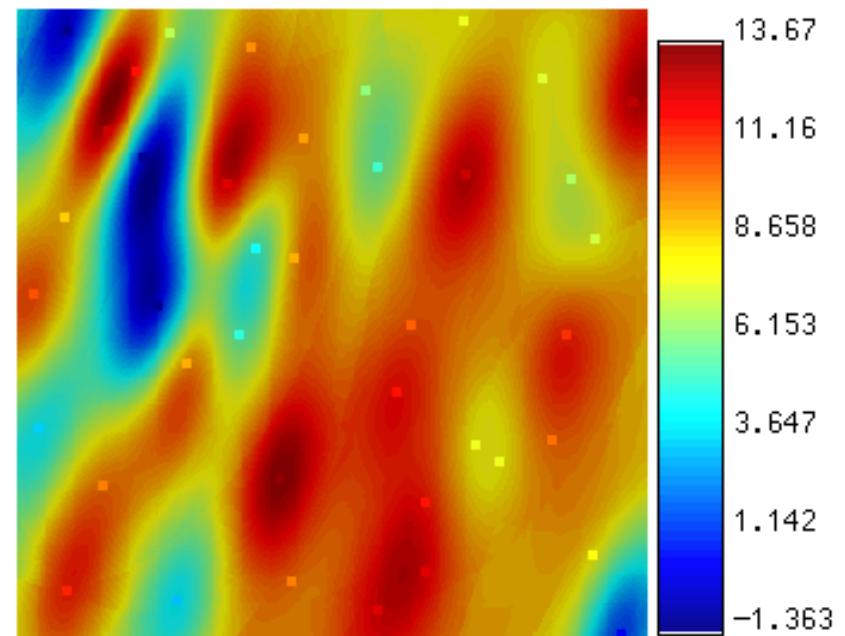
Ordinary kriging map
40 measurements

CROSSVALIDATION
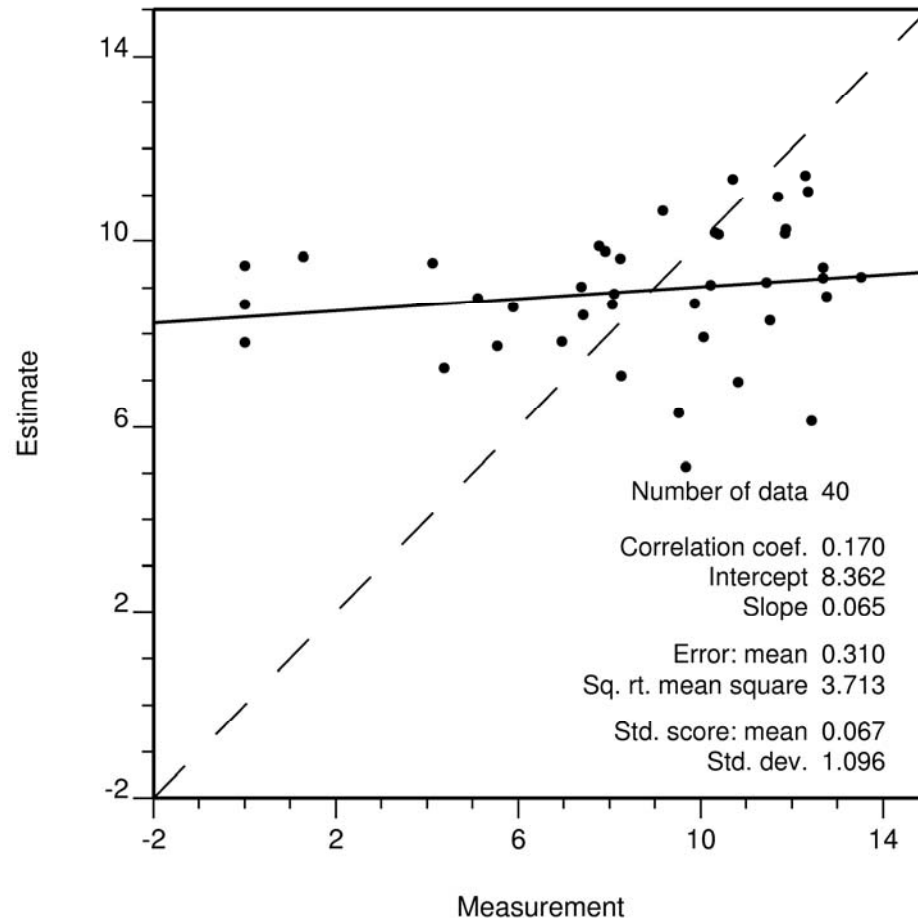
Estimated values are close to being pure noise, with estimates for low measurements above the main diagonal and conversely for the highest data values. This a form of smoothing.

SEMIVARIOGRAMS OF ORDINARY KRIGING MAP
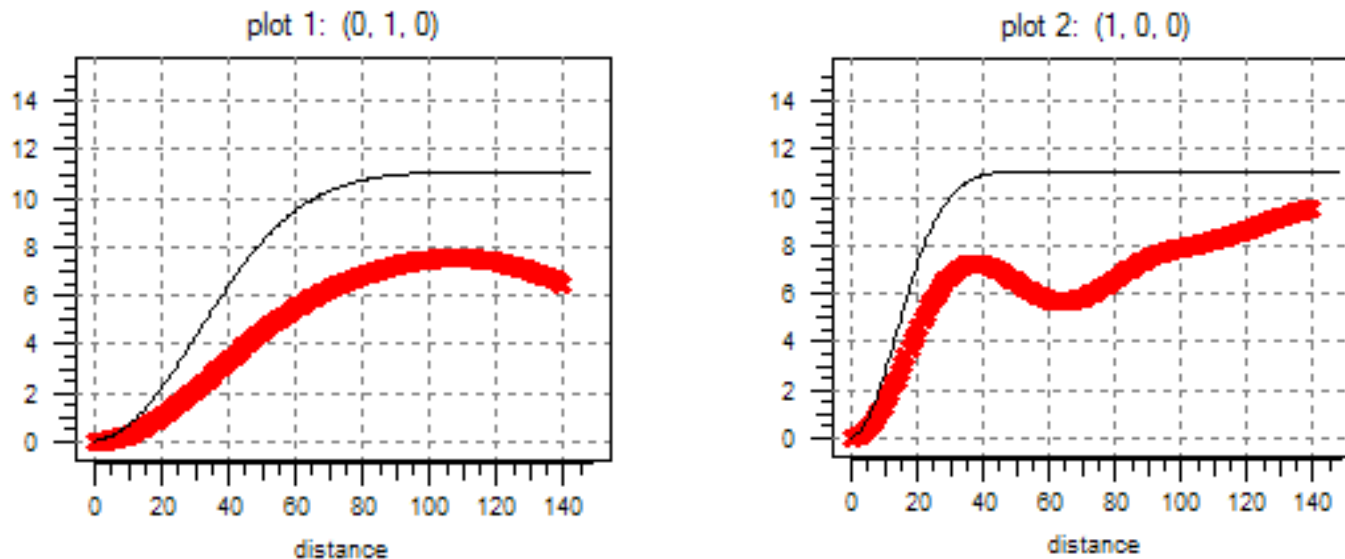
There is wide disagreement between the actual semivariogram and that of the ordinary kriging map. This is a consequence of smoothing.
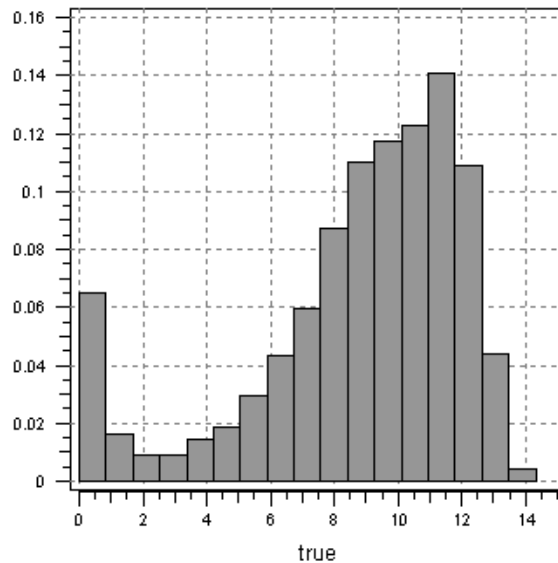
258
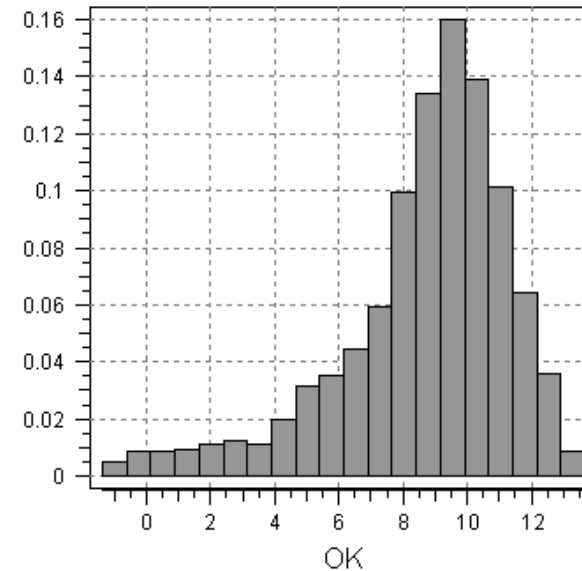
HISTOGRAMS

The ordinary kriging histogram is not bimodal like the histogram of the exhaustive sample, which is also a consequence of smoothing.

CUMULATIVE DISTRIBUTIONS
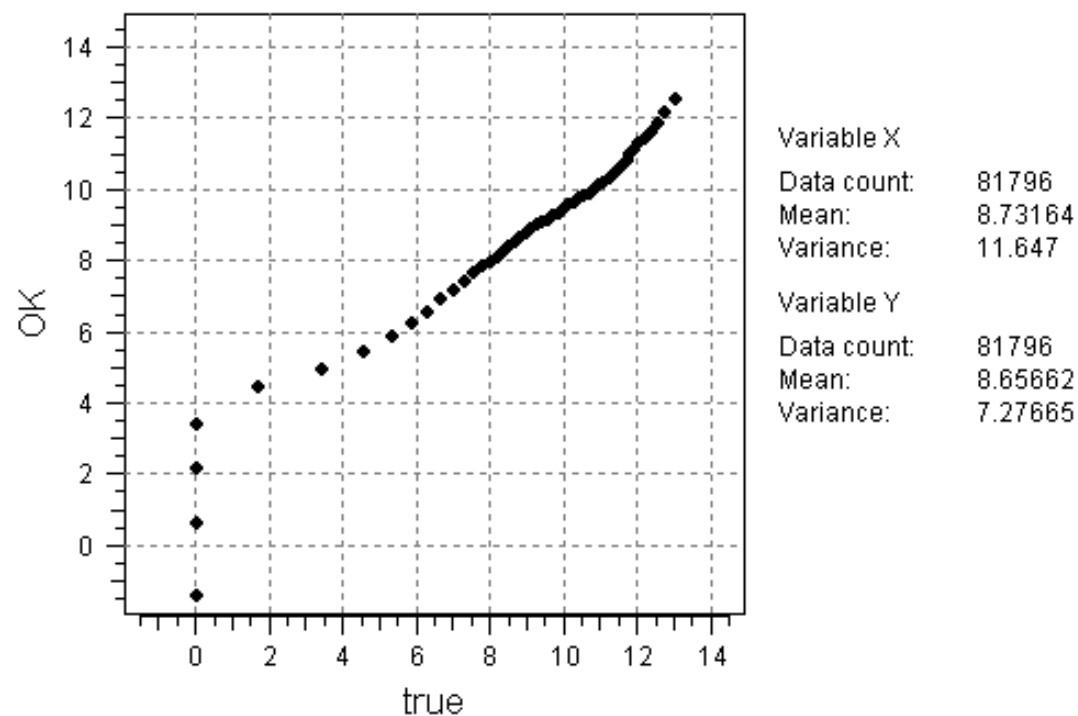
Discrepancy in the proportion of zeros results in clearly different distributions.

ERRORS ACCORDING TO EXHAUSTIVE SAMPLE
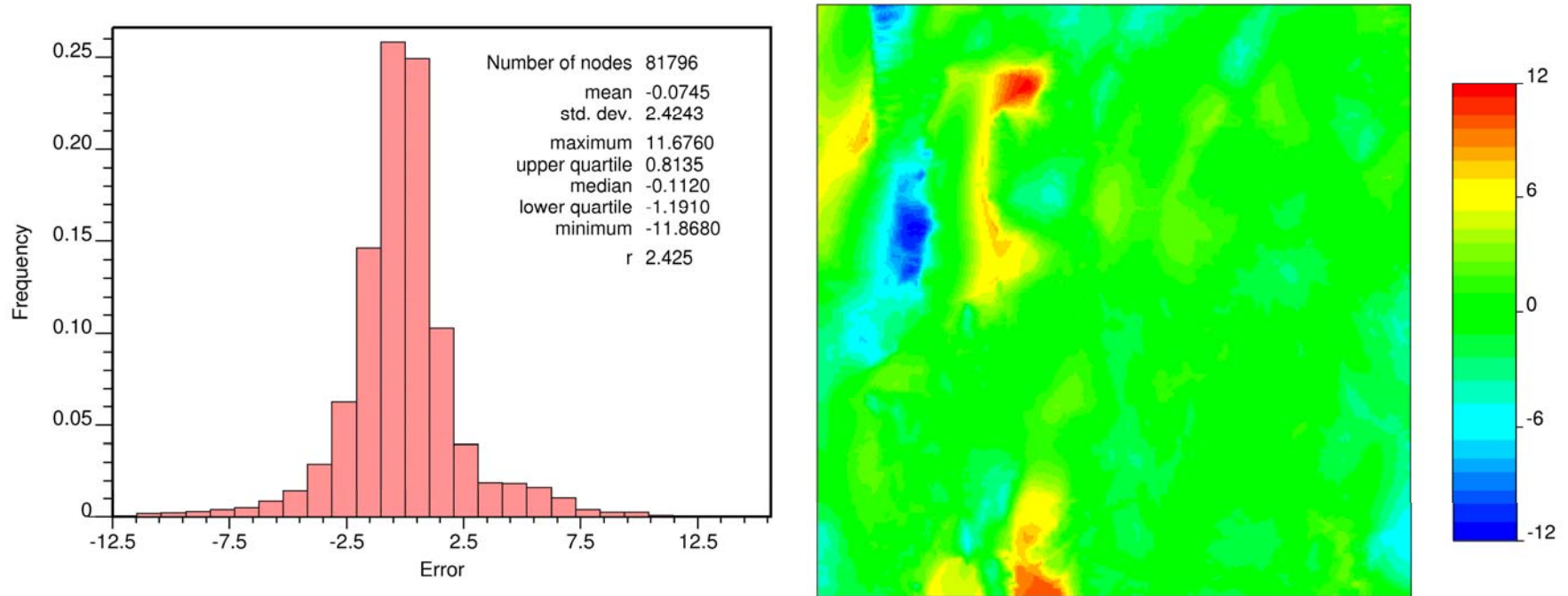
In agreement with the global unbiasedness of kriging, the errors are symmetric around a value approximately equal to zero.

SEMIVARIOGRAM OF ERRORS ACCORDING TO EXHAUSTIVE SAMPLE
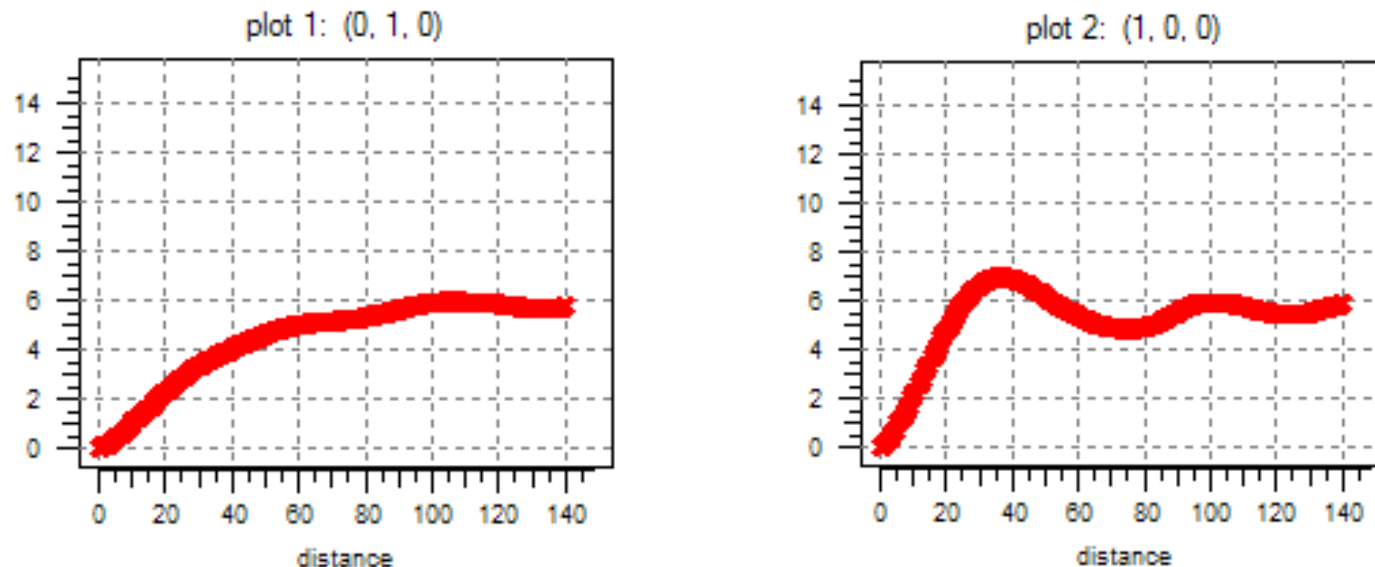
Errors are spatially correlated, with a range remarkably similar to that of the sample, but with a significantly smaller sill.

A VIEW AT ERRORS AND KRIGING STANDARD DEVIATIONS
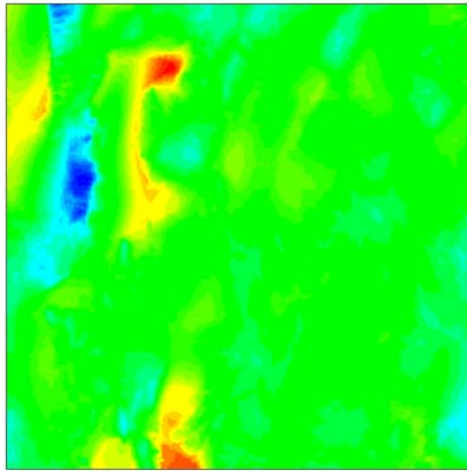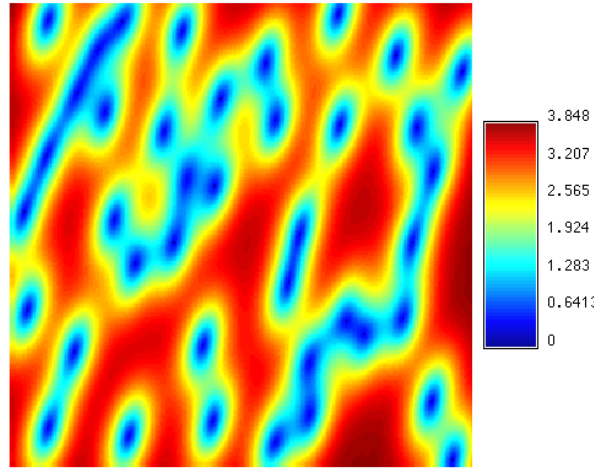
Errors

Kriging standard deviation
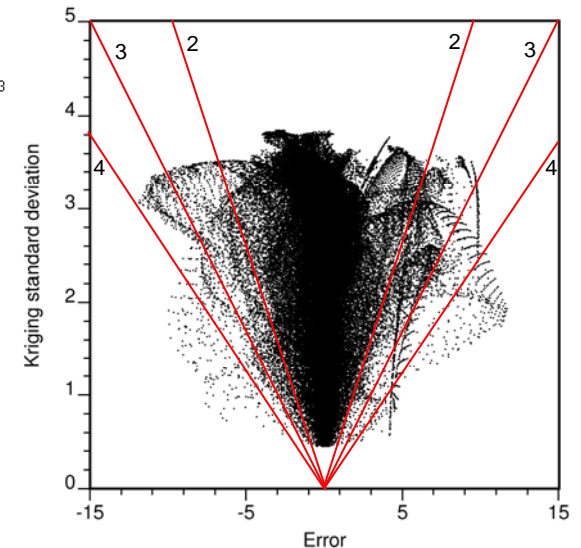
Most kriging standard deviations are no more than twice the error, but there is no correlation, nor the type of assurance provided by the Chebyshev's theorem.

REMEDIAL ALTERNATIVES

There have been efforts along two different lines to eliminate smoothing:

- Compensation procedures
- Simulation

Success of compensation procedures has been limited.

By far the bulk of the effort and results have been in the generation of equiprobable realizations using simulation.

When simulation is used, uncertainty is modeled by generating several realizations. The wider the discrepancies, the larger the uncertainty.

SIMULATION METHODS

The most common methods are:

- Turning bands method
- Lower-upper triangular method
- Truncated plurigaussian methods
- Object oriented methods
- Sequential Gaussian simulation
- Simulated annealing
- Filter simulation

   When a realization by any method reproduces the value and location of all observations, the procedure is said to be a conditional simulation. Otherwise, it is an unconditional simulation.

15. SEQUENTIAL GAUSSIAN SIMULATION

BACKGROUND

Sequential Gaussian simulation is a stochastic method that is basically multigaussian kriging with feedback.

Let $Z(\mathbf{s})$ be a subset of N variates of a random function and let $z(\mathbf{s})$ be a sample of size n. As a consequence of Bayes's theorem, one can break a bivariate cumulative frequency distribution function into the product of two univariate cumulative frequency distribution functions:

$$F(\mathbf{s}_1,\mathbf{s}_2,t_1,t_2 \mid n) = F(\mathbf{s}_1,t_1 \mid n) \cdot F(\mathbf{s}_2,t_2 \mid n+1)$$

where $F(\mathbf{s}_2,t_2 \mid n+1)$ is the probability $\Pr(Z(\mathbf{s}_2) \le t_2)$, conditional to the original sample plus a value $z(\mathbf{s}_1)$ drawn from the distribution $F(\mathbf{s}_1,t_1 \mid n)$.

EXTENSION OF THE DECOMPOSITION

By extension, the decomposition of a bivariate distribution can be generalized to any number of variates.

$$F\left(\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N, t_1, t_2, \ldots, t_N \mid n\right) =$$
$$F\left(\mathbf{s}_1, t_1 \mid n\right) \cdot F\left(\mathbf{s}_2, t_2 \mid n+1\right) \cdot \ldots \cdot F\left(\mathbf{s}_N, t_N \mid n+N-1\right)$$

Sequential Gaussian simulation is a practical implementation for the generation of numerical realizations of the joint distribution $F\left(\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N, t_1, t_2, \ldots, t_N \mid n\right)$.

TEN-STEP APPROACH (1)

To generate a partial realization of a random function at the nodes of a regular grid, do the following:

1. In case the sample is not univariate normal, transform the data into normal scores. SGeMS can do it, going to <Algorithms> then <Utilities> then <transf>.

2. Using standard semivariogram modeling techniques, find the best semivariogram model for the transformed data.

3. Pick nodes at random and stack them all in a visitation queue.

4. For faster execution, move the data to the closest node. In case two or more observations land in the same node, implement a criterion to end up with one value per node, such as averaging or discarding.

TEN-STEP APPROACH (2)

5. From the node sequence queue, draw the next location $\mathbf{s}_i$ due for the calculation of a simulated value. Employ multigaussian kriging to find the estimate $z^*(\mathbf{s}_i)$ and the corresponding estimation variance $\sigma^2(\mathbf{s}_i)$ using an expanded sample comprising all n original data plus all values that may have been already simulated.

6. Use $z^*(\mathbf{s}_i)$ and $\sigma^2(\mathbf{s}_i)$ to define a normal distribution with mean $z^*(\mathbf{s}_i)$ and variance $\sigma^2(\mathbf{s}_i)$, $N\!\left(z^*(\mathbf{s}_i), \sigma^2(\mathbf{s}_i)\right)$.
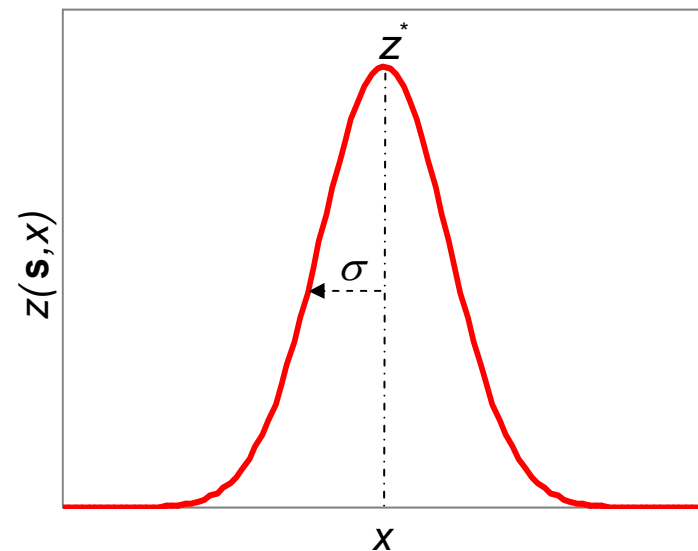
TEN-STEP APPROACH (3)

7. Draw at random a value, $z_s(\mathbf{s}_i)$, from $N\left(z^*(\mathbf{s}_i), \sigma^2(\mathbf{s}_i)\right)$. The number $z_s(\mathbf{s}_i)$ is the simulated value at location $\mathbf{s}_i$.

8. Add $z_s(\mathbf{s}_i)$ to the expanded sample.

9. If $\mathbf{s}_i$ is not the last node in the queue, then go back to Step 5.

10. In case Step 1 was necessary, backtransform all the values in the partial realization to the original sampling space.
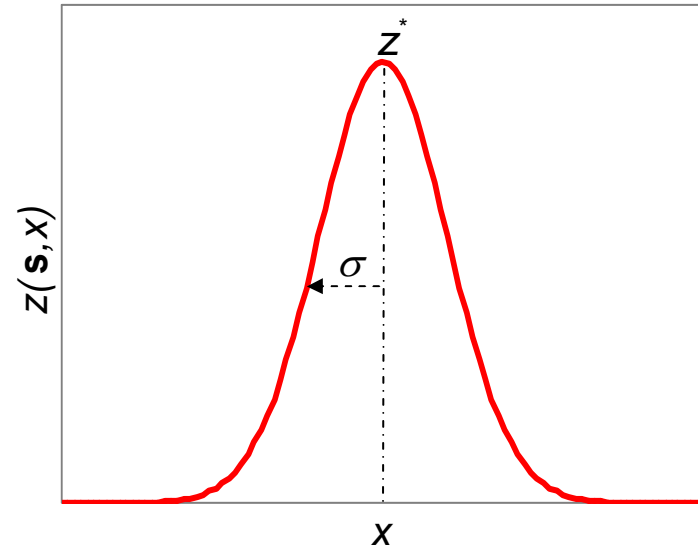
EXERCISE

Use SGeMS and the Ely West sample of size 40 (page 188) to do the following:

A. Generate 100 realizations by sequential Gaussian simulation; compare the first one to the exhaustive sample.

B. Examine the semivariograms for the first realization along the north-south and east-west directions and compare them to the exhaustive semivariogram.

C. Compare the histograms for the exhaustive sample and the first realization.

D. Prepare a Q-Q plot to compare the same distributions.

E. Use the exhaustive sample to obtain the errors for the first realization. Map them and study the semivariogram.

F. Investigate errors in the first realization and the standard deviation that results from all realizations at every node.

A. SGeMS CODE FOR GENERATING 100 SEQUENTIAL GAUSSIAN SIMULATIONS
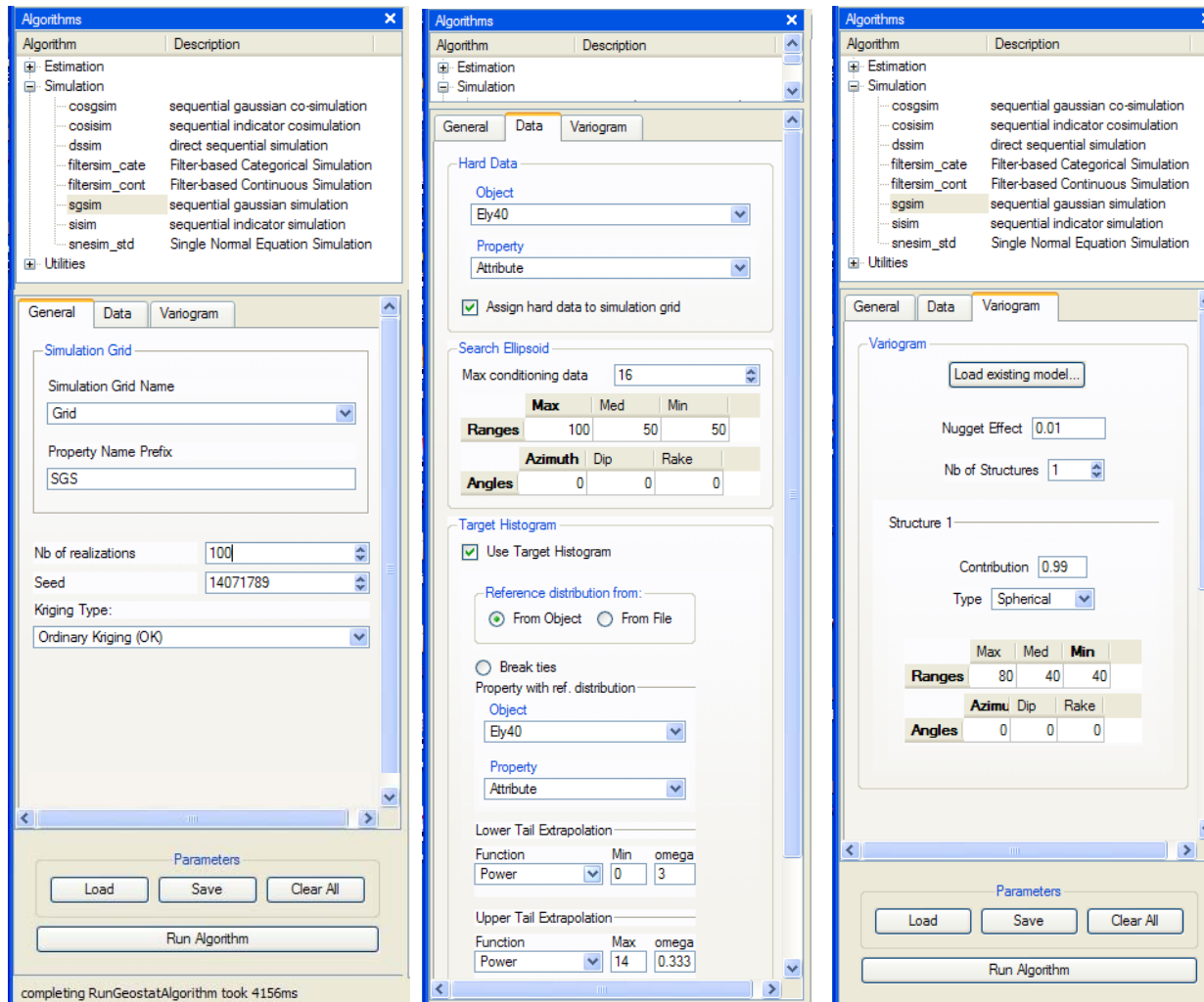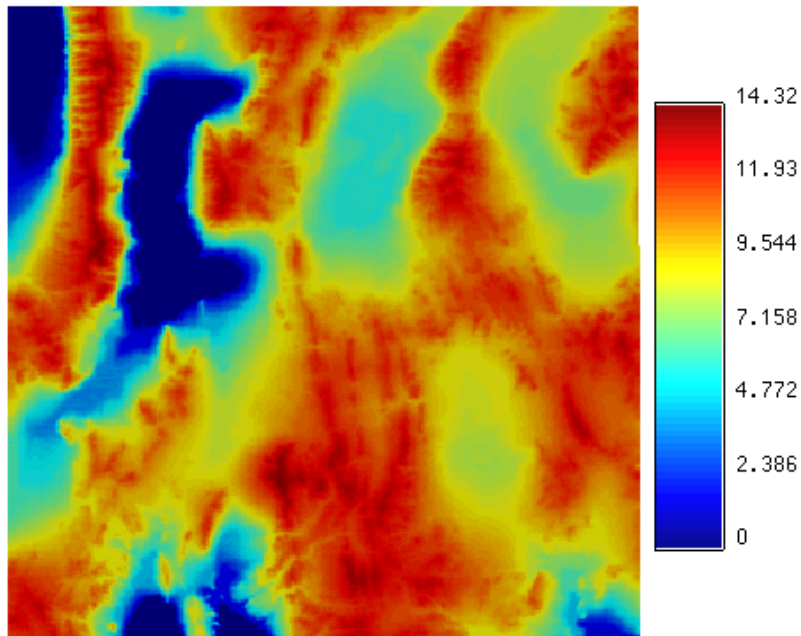
- Note that the attribute is the data, not the normal scores.
- However, the semivariogram is for the normal scores, not for the attribute.
- No backtrans-formation is necessary.

TWO REALIZATIONS FOR ELY WEST
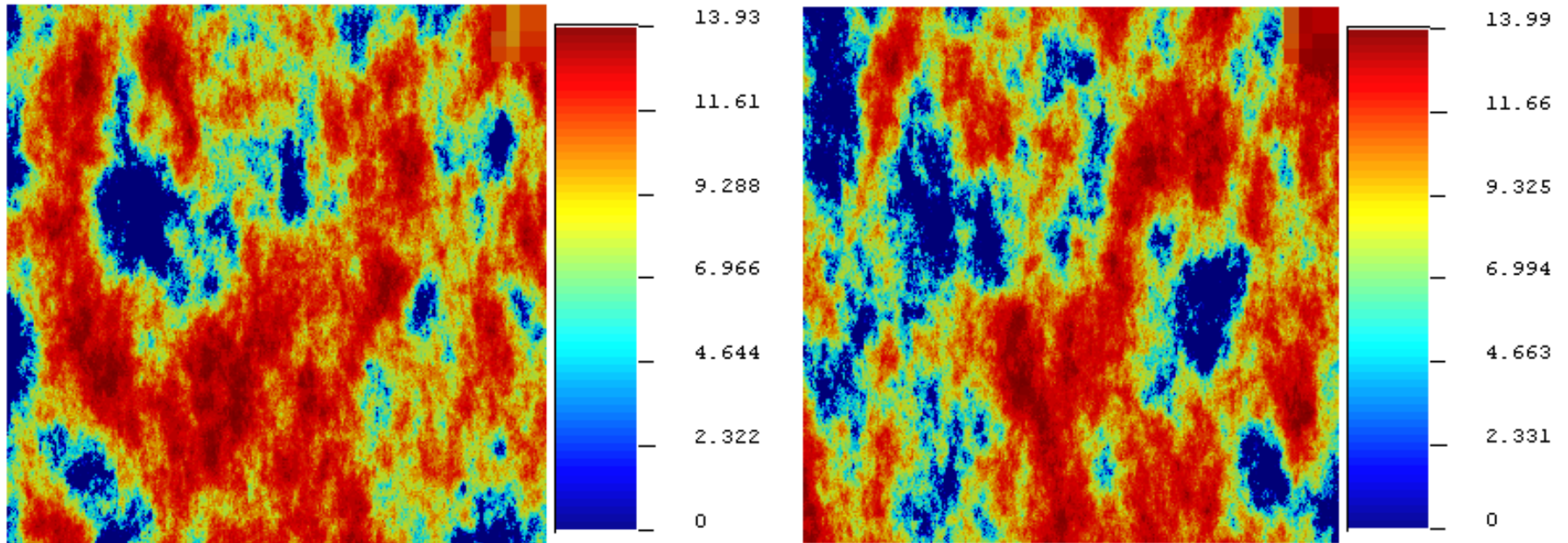
These are conditional realizations honoring the same 40 data. Thus, although hard to note, the values of both realizations at those 40 nodes are the same.

BACK TO THE EXHAUSTIVE CASE

Exhaustive sample
81,796 measurements
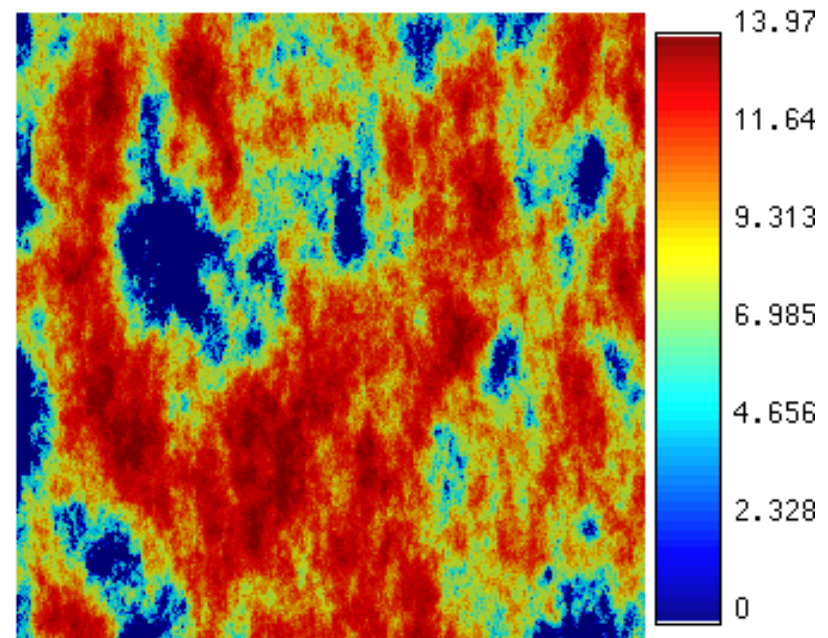
First realization
40 measurements

B. FIRST REALIZATION SEMIVARIOGRAMS
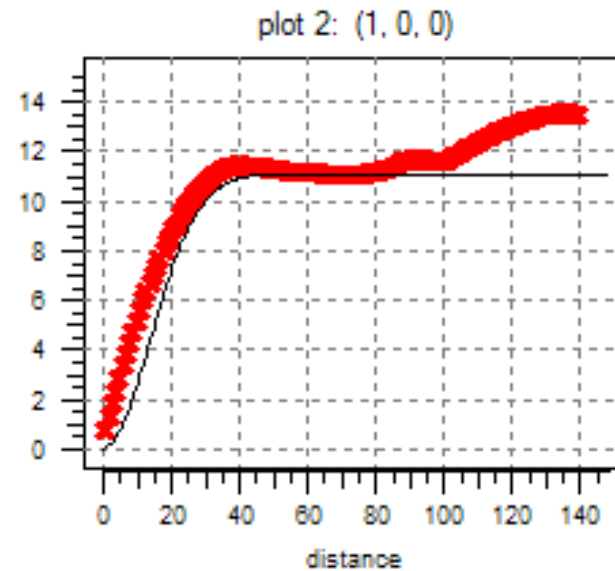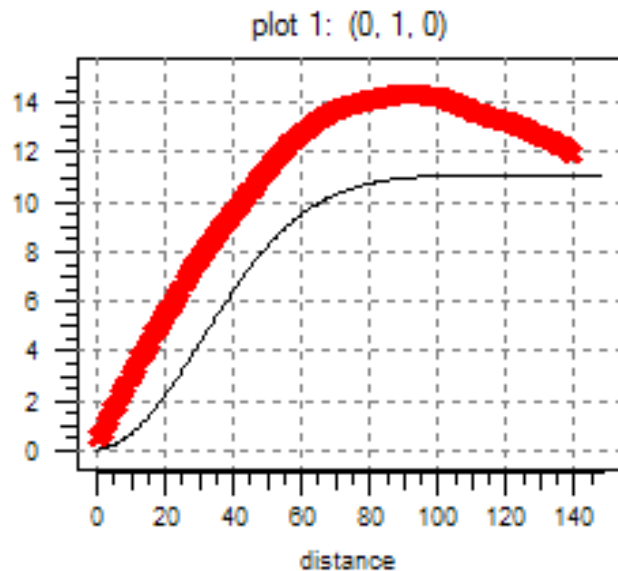
Agreement is not perfect but is significantly better than in the case of kriging.

C. HISTOGRAMS
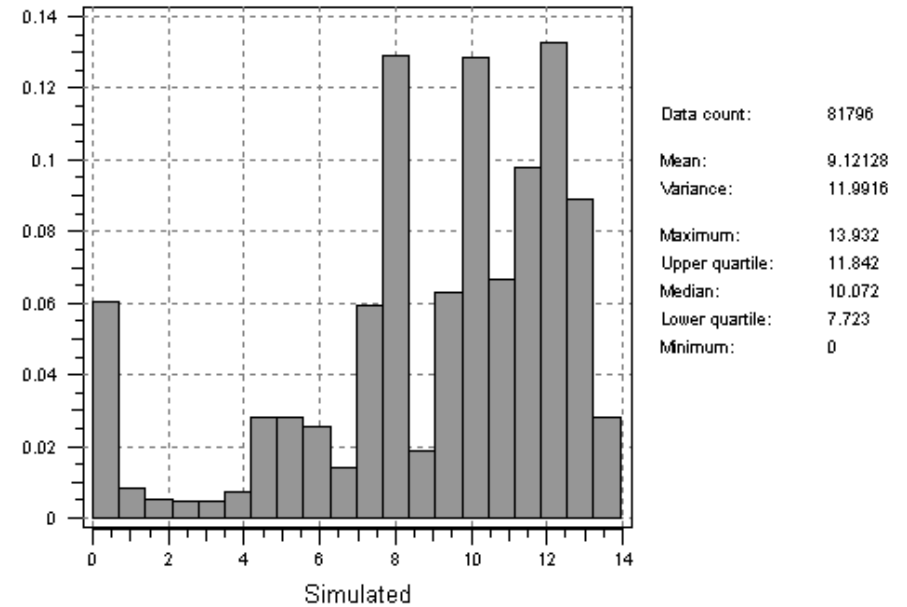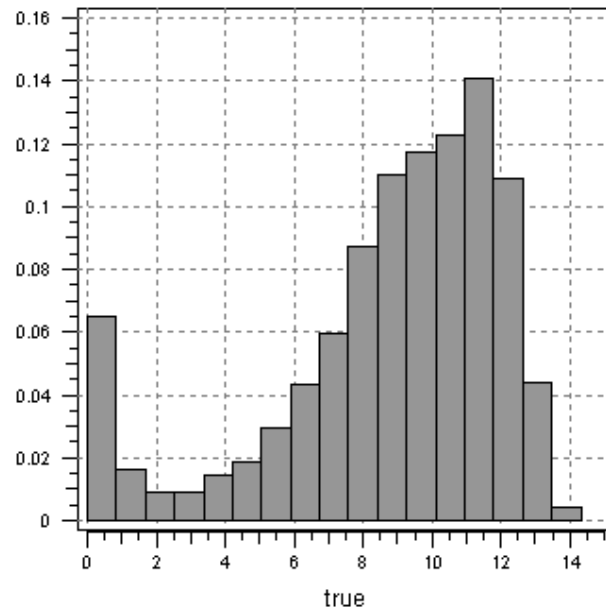
The realization follows the sample histogram, not that of the exhaustive sample.

D. CUMULATIVE DISTRIBUTIONS
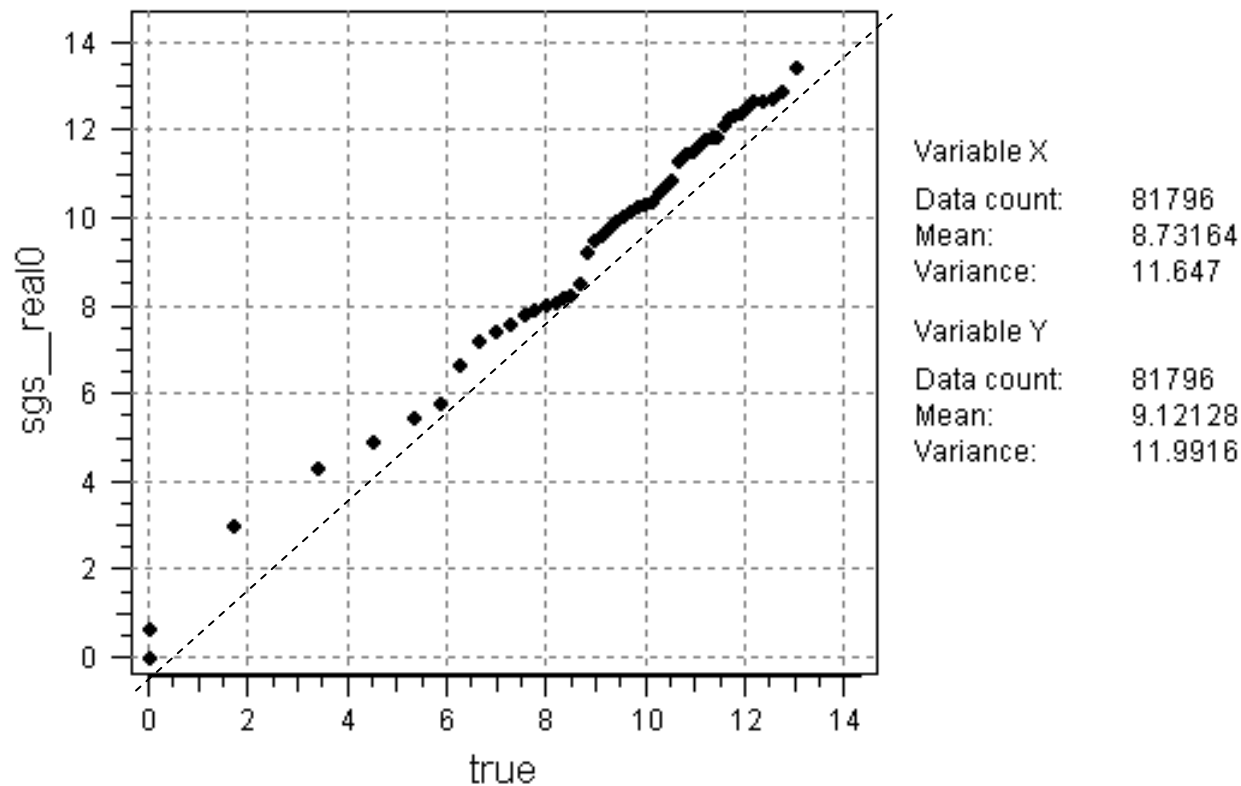
According to the Q-Q plot, agreement between distributions is also better than in the case of ordinary kriging.

E. ERRORS ACCORDING TO EXHAUSTIVE SAMPLE
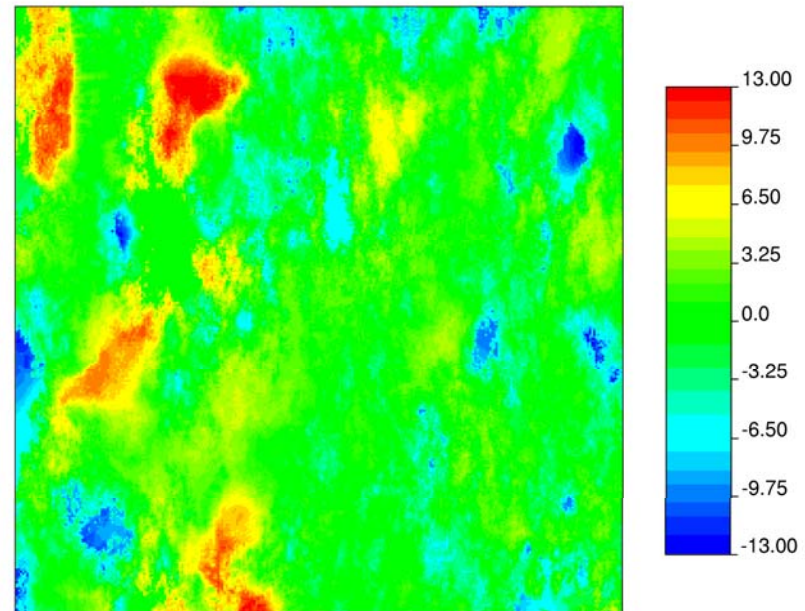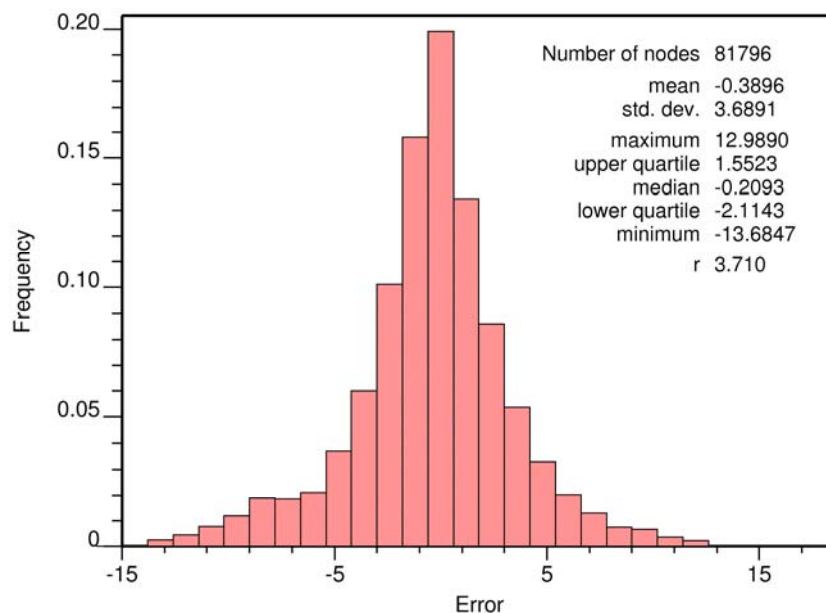
Errors are unbiased but are larger than those for kriging.

SEMIVARIOGRAM FOR ERROR IN FIRST REALIZATION
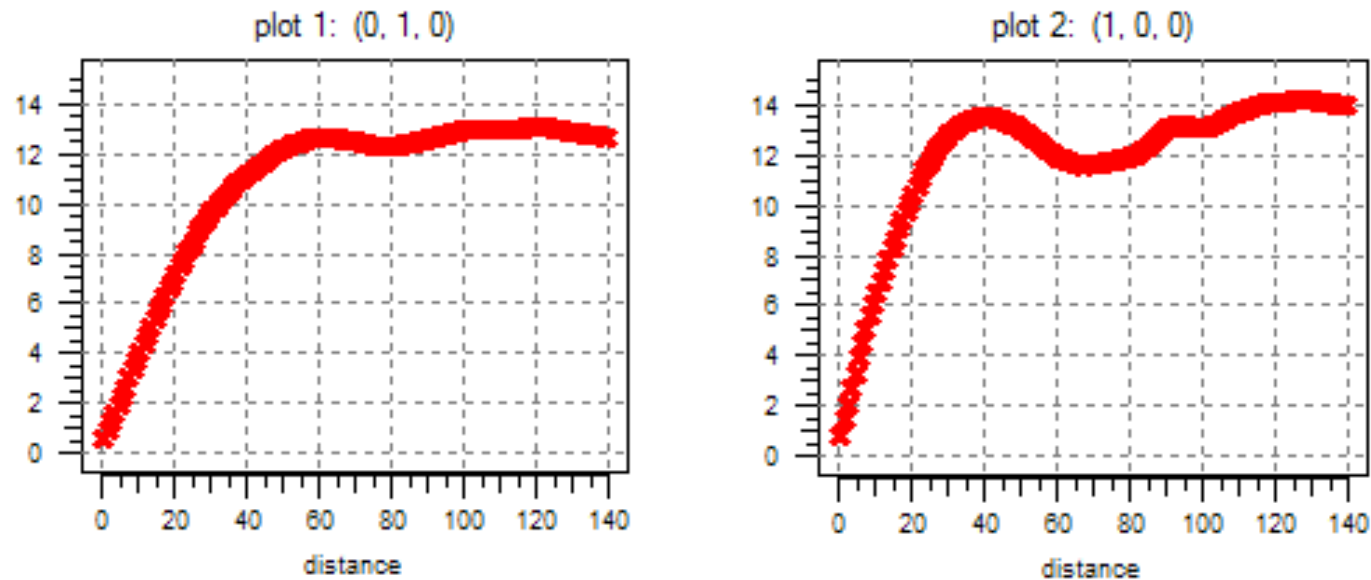
The error semivariogram resembles that of the attribute.

F. A LOOK AT FIRST REALIZATION ERRORS AND NODAL STANDARD DEVIATIONS
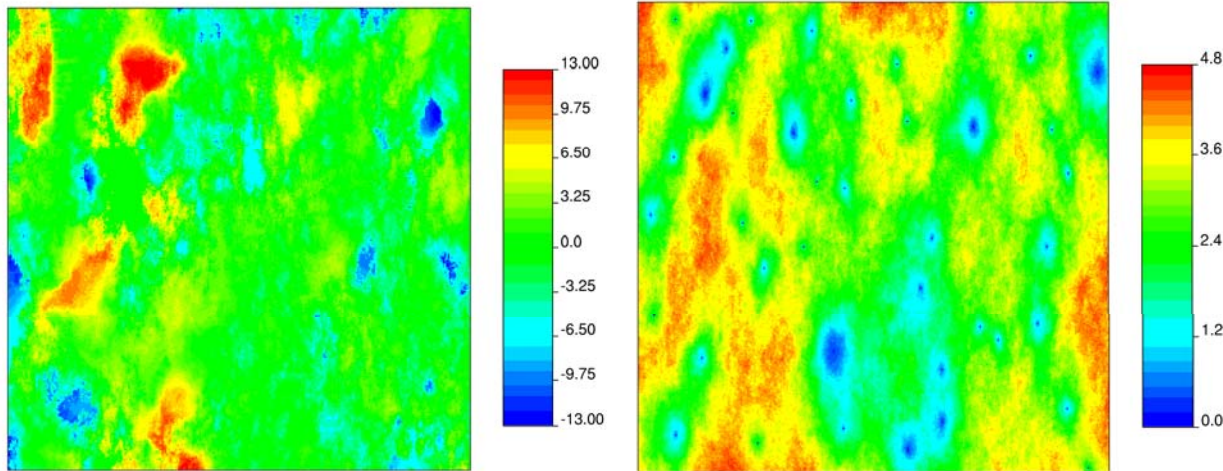
Errors

Nodal standard deviation

Given a node, the nodal standard deviation is the standard deviation of the values from all realizations at that node.

  Larger standard deviations are associated with large errors, but there is no correlation.
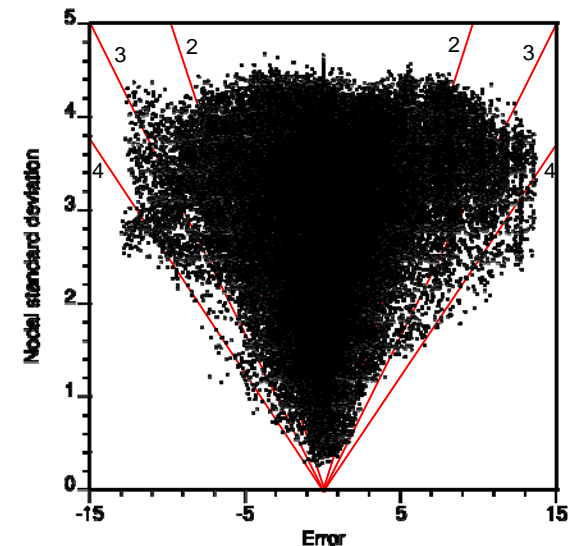
16. SIMULATED ANNEALING

THE GENERAL IDEA

Simulated annealing was one of the first methods for the modeling of spatially correlated attributes that moved away from generating values for uninformed nodes by combining the sample values.

Simulated annealing started as an approach to model the cooling of a piece of hot metal, but it is now more widely used as an optimization tool instead.

In geostatistics, simulated annealing is used to move around node values in order to have a realization with some property as close as possible to a target function, which commonly is the semivariogram.

Its main feature is the lack of any distributional assumptions.

"COOLING" A GRID

A simulation grid is populated by placing each sample value at the collocated or closest node. The unsampled nodes are given values randomly taken from a probability distribution, typically that of the sample.

By so doing, the realizations honor the data and the selected histogram, but in a typical situation, unless the sample is large relative to the grid size, the grid is totally chaotic and lacking spatial correlation.

The aim of simulated annealing is to transform the grid to gain additional properties, customarily to be spatially correlated by following a semivariogram model.

OBJECTIVE FUNCTION

When the target function is a semivariogram model $\gamma(\mathbf{h})$, the objective function G is:

$$G = \sum_{\mathbf{h}} \frac{[\gamma_G(\mathbf{h}) - \gamma(\mathbf{h})]^2}{\gamma(\mathbf{h})^2}$$

where $\gamma_G(\mathbf{h})$ is the semivariogram of the realization.

Reduction in the value of G is achieved through swapping pairs of values at nodes which are not sampling locations.

Key to a successful process is not to reject 100 percent of the exchanges of nodal values that do not result in an improvement of the objective function G.

The cooling schedule controls the proportion of unfavorable swaps as a function of total swaps attempted.

COMPONENTS OF AN ANNEALING SCHEDULE

$$\Pr[accept\ swap] = \begin{cases} 1, if\ G_{new} \leq G_{old} \\ e^{\frac{G_{old}-G_{new}}{t}}, \text{otherwise.} \end{cases}$$

t is the temperature, starting at t_0;

λ temperature reduction factor;

K_a is the maximum number of all swaps for a given temperature, after which it is reduced by λ;

K is the maximum number of accepted swaps for a given temperature, after which it is reduced by λ;

S is the maximum number of times for reaching K_a;

G_{min} is the acceptable value for the objective function.

ANNEALING IN ACTION
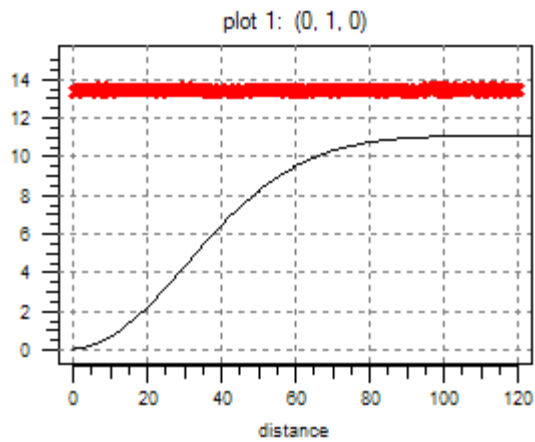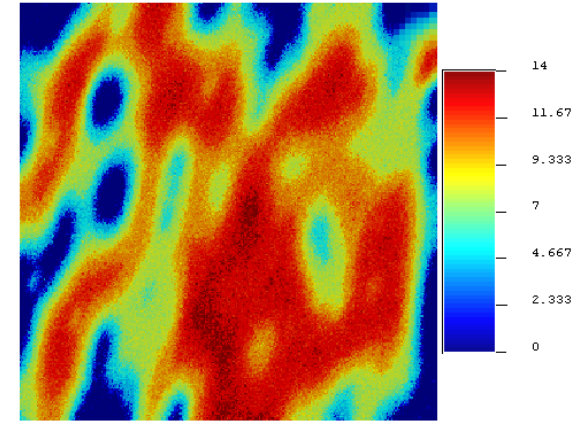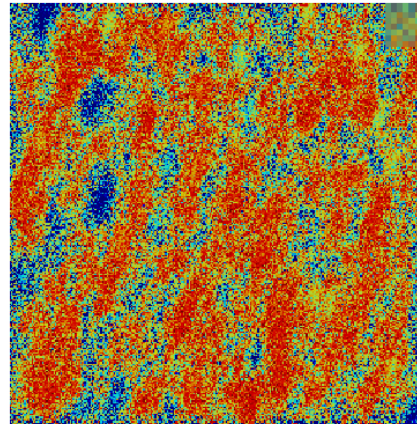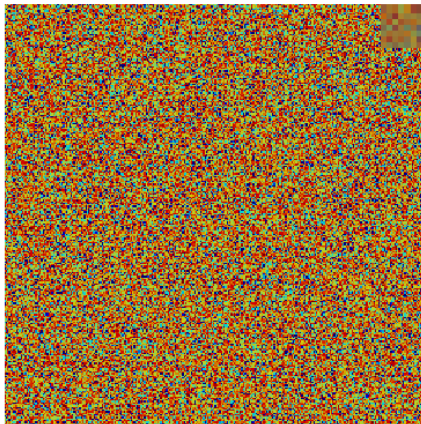
G=1
Swaps=0
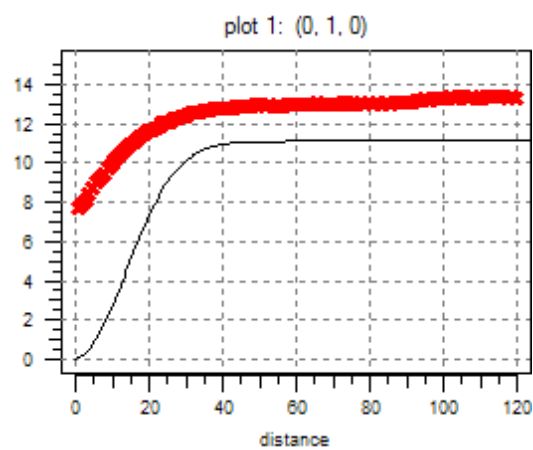
G=0.2
Swaps=2,640,088
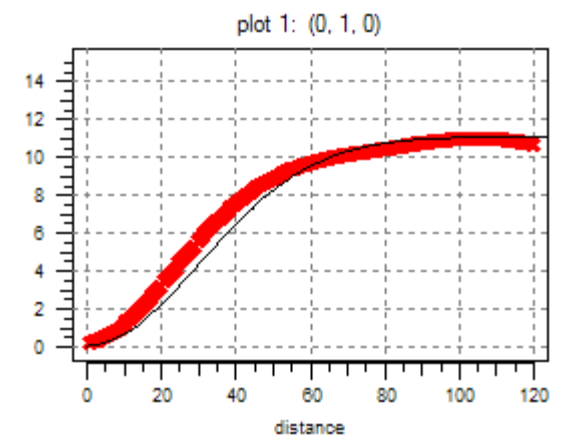
G=0.0001
Swaps=10,895,372

FEATURES

- No computational power at all is used to calculate node values.

- As all nodal values are drawn from the sample distribution, its representativity is more critical than for the other simulation methods.

- All efforts are devoted to moving around node values.

- Given a sample, processing time is in general the longest among simulation methods, which increases more steeply than linearly with the number of grid nodes.

- Despite its novel computational approach, simulated annealing continues to rely on a sample and global properties that can be derived from the sample, most commonly the semivariogram.

- The most salient property is independence from any distributional assumptions.

REPETITION OF PAGE 272 EXERCISE NOW FOR SIMULATED ANNEALING

The method is not part of SGeMS, but it is in GSLIB.

```
                    Parameters for SASIM (1992)
                    ***************************

START OF PARAMETERS:
s40.dat                                   \conditioning data (if any)
1   2   0   4                             \columns: x,y,z,vr
-1.0e21    1.0e21                          \data trimming limits
0                                         \0=non parametric; 1=Gaussian
s40.dat                                   \non parametric distribution
4   0                                     \columns: vr,wt
0   14.0                                  \minimum and maximum data values
1      1.0                                \lower tail option and parameter
4      2.0                                \upper tail option and parameter
sasim.out                                 \output File for simulation
sasim.var                                 \output File for variogram
3     500000                                \debug level, reporting interval
sasim.dbg                                 \output file for debugging
1                                         \annealing schedule? (0=auto)
1.0   0.12   4000000   400000   5   0.0001   \manual schedule: t0,lambda,ka,k,e,Omin
2                                         \1 or 2 part objective function
112063                                    \random number seed
100                                       \number of simulations
286 0 1                                   \nx,xmn,xsiz
286 0 1                                   \ny,ymn,ysiz
1    0.0    1.0                            \nz,zmn,zsiz
150                                       \max lags for conditioning
1   0.1   0                               \nst, nugget, (1=renormalize)
3   45.0   13.0                           \it,aa,cc:        STRUCTURE 1
15.0  0.0  0.0  0.45  0.0                 \ang1,ang2,ang3,anis1,anis2:
```

A. TWO REALIZATIONS FOR ELY WEST
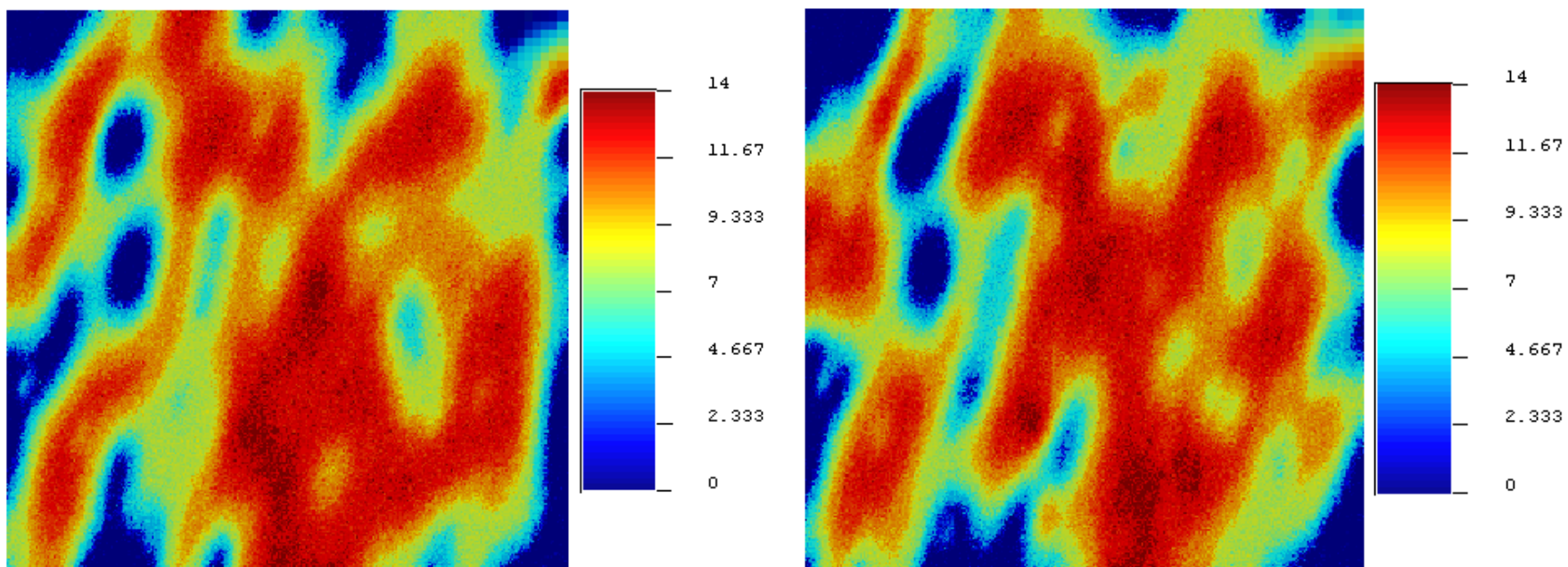
BACK TO THE EXHAUSTIVE CASE

Exhaustive sample
81,796 measurements
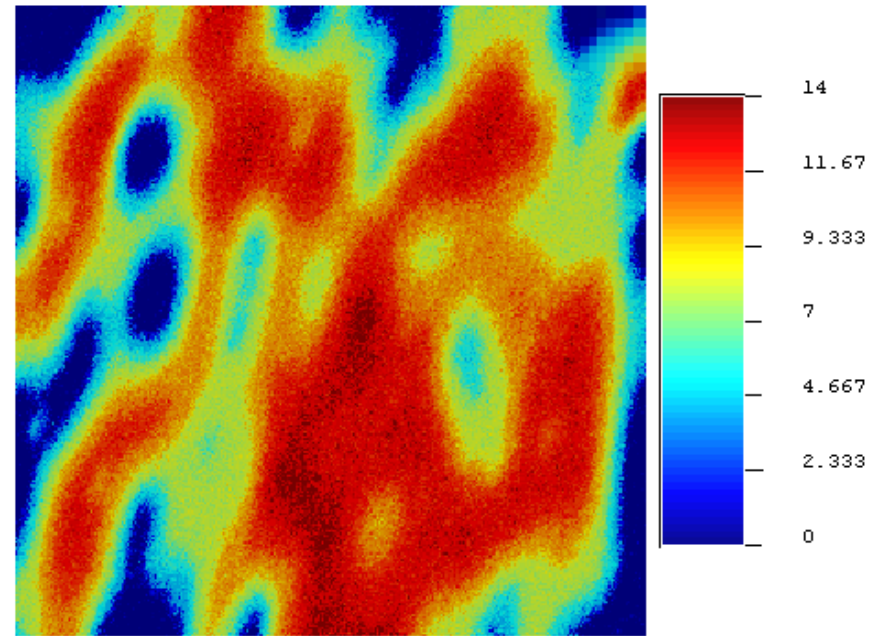
First realization
40 measurements

B. SEMIVARIOGRAMS FOR
THE FIRST REALIZATION
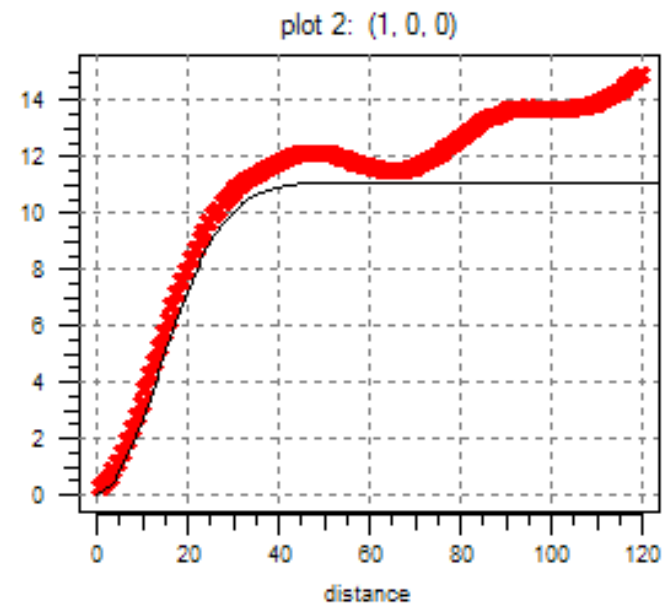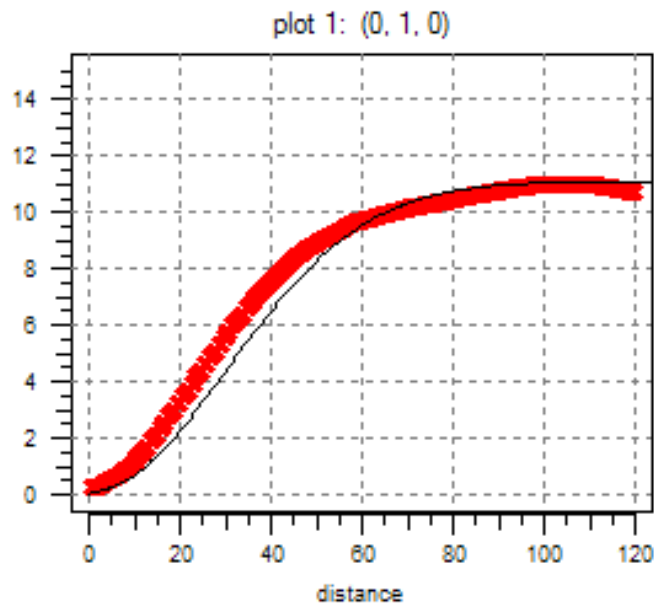
Semivariogram reproduction is good.

C. HISTOGRAMS
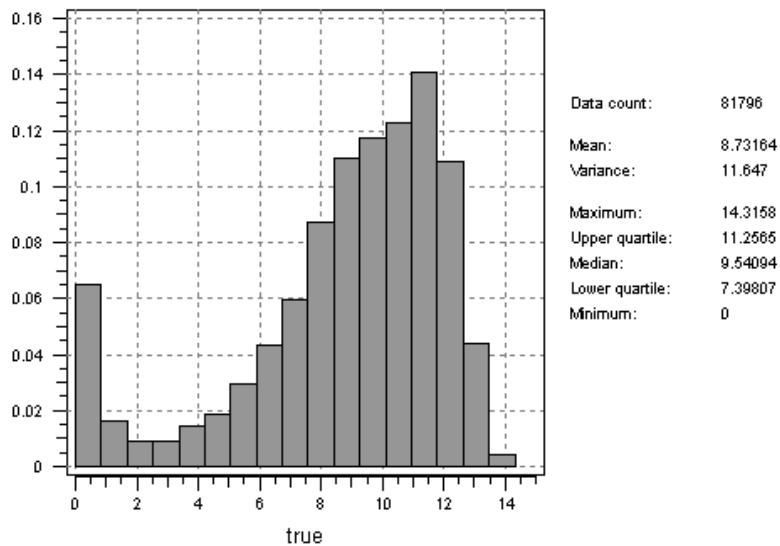
Exhaustive sample
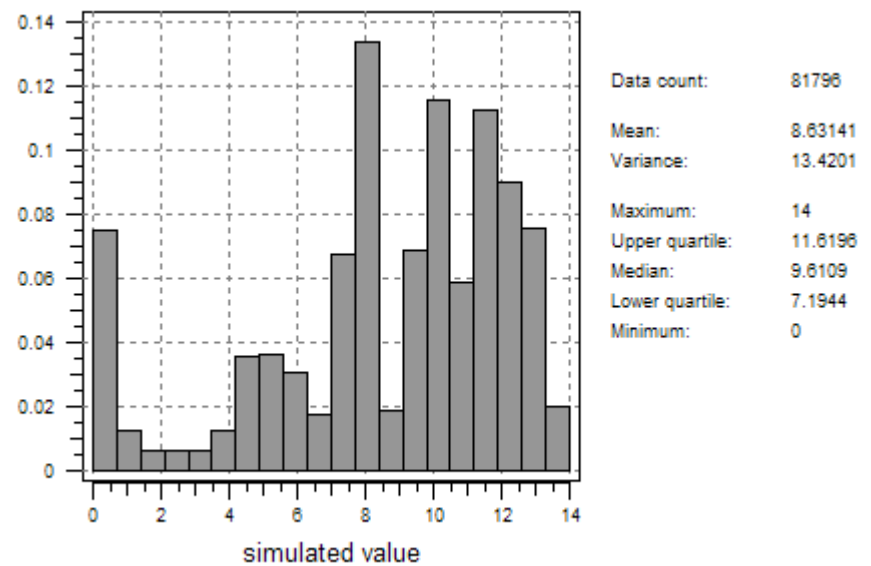
First realization

There is also good reproduction of the sample histogram.

D. CUMULATIVE DISTRIBUTIONS
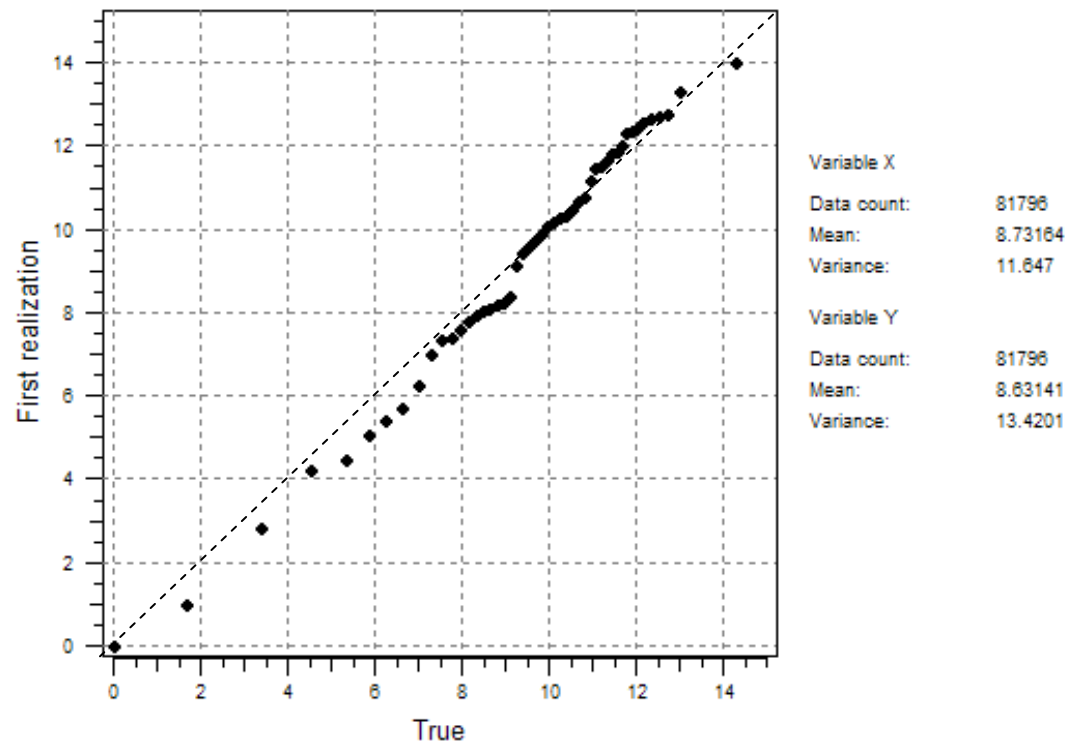
Agreement between the distributions of realizations and the population is limited only by the unbiasness of the empirical sample.

E. ERRORS IN FIRST REALIZATION ACCORDING TO EXHAUSTIVE SAMPLE
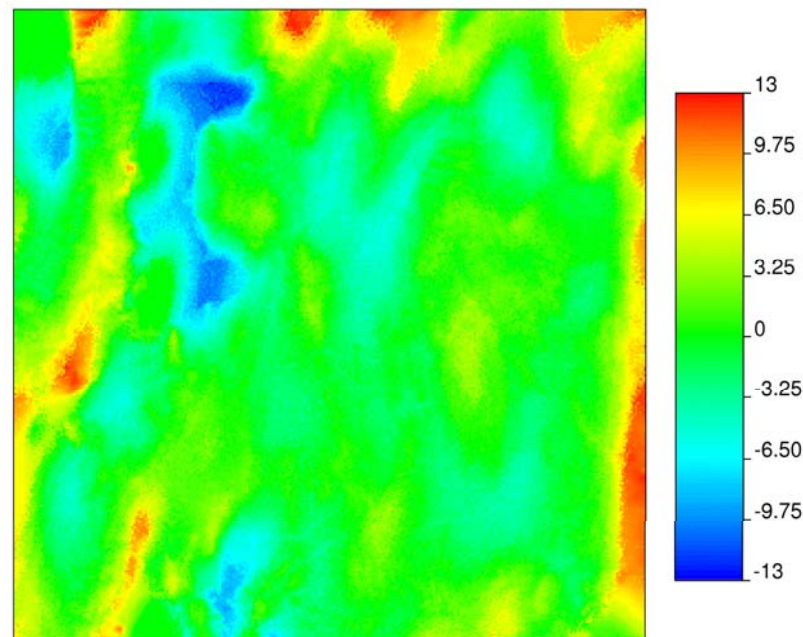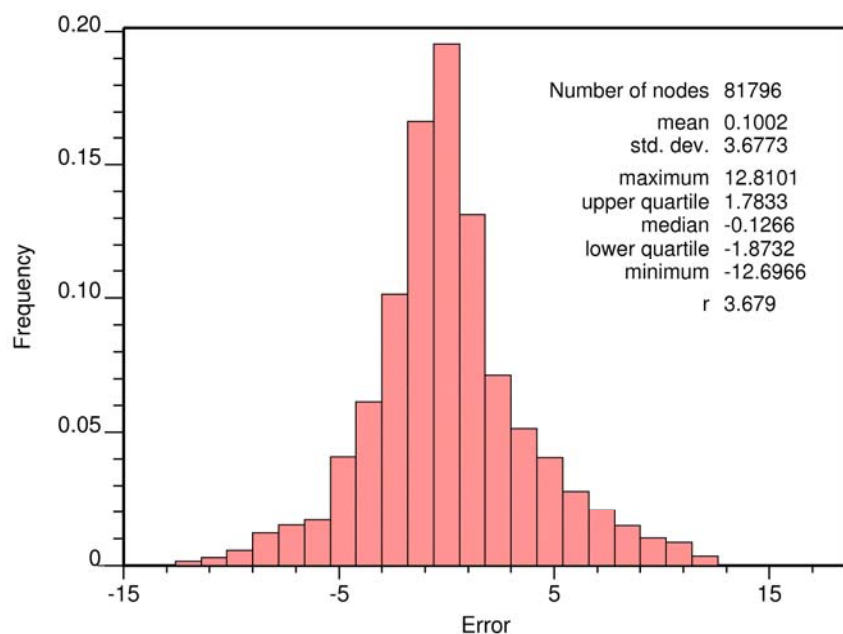
Errors are comparable to those of sequential Gaussian simulation.

ERROR SEMIVARIOGRAMS
FOR FIRST REALIZATION
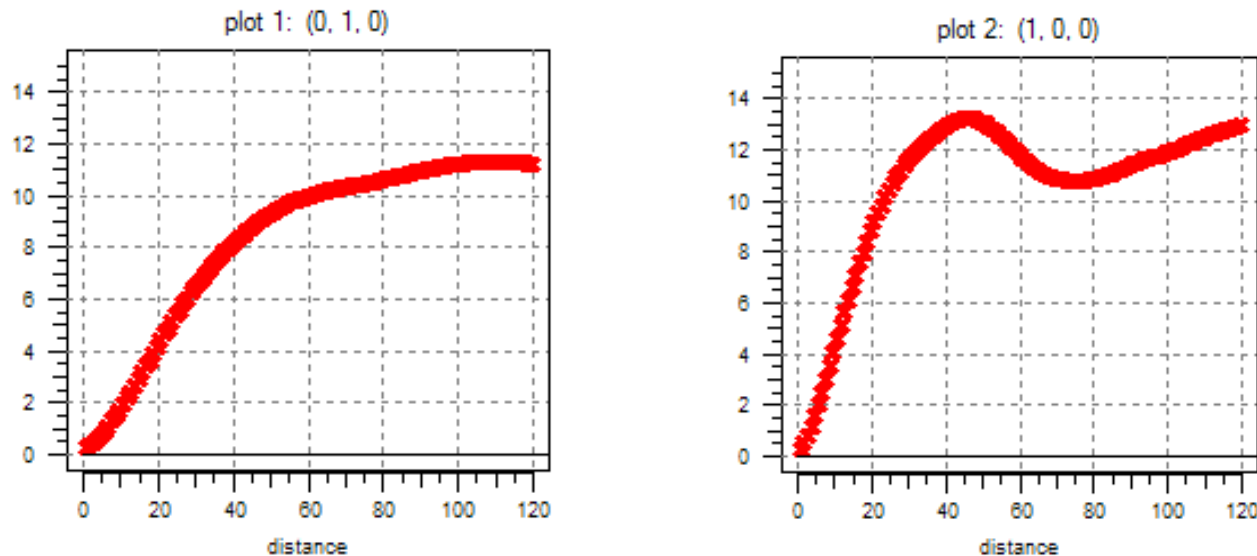
Errors have a semivariogram resembling the semivariogram of the attribute.

F. TRYING TO RELATE ERRORS AND NODAL STANDARD DEVIATIONS

The tendency of errors to stay within four nodal standard deviations is still true, but it is less pronounced than for kriging and sequential Gaussian simulation.
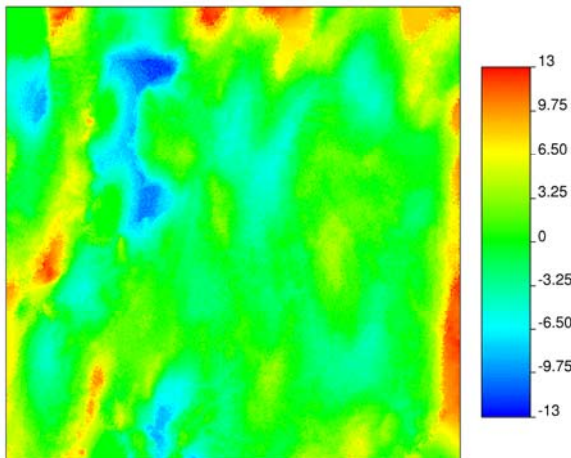
Errors in first realization
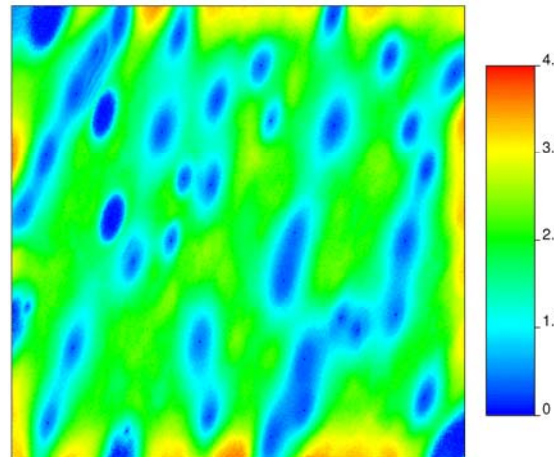
Nodal standard deviation
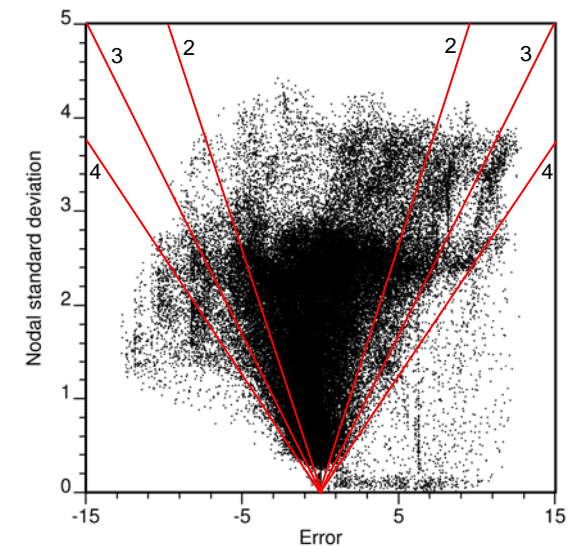
17. FILTER SIMULATION

CLASSICAL GEOSTATISTICS

- Until recently, geostatistics had evolved to provide modeling tools relying on methods typical of stochastic processing.

- Central to past developments is the heavy use of moments up to order 2, particularly the semivariogram or its equivalent the covariance.

MULTIPLE POINT STATISTICS

- The first attempts to expand classical geostatistics aimed to calculate statistics involving three or more points.

- These efforts failed because of the high demands in terms of sample size required for the estimation of higher order moments.

MULTIPLE POINT RELATIONSHIPS

- However, the idea of considering relations among several points has stayed.

- The concept of moment has been abandoned in favor of techniques related more to image analysis and pattern recognition, such as filtering and cluster analysis.

  Filter simulation is a member of the multipoint statistics family of simulation methods. It is based on pattern recognition and does not require the semivariogram at all, even though one can still estimate it for purpose of evaluation or comparison.

MULTIPOINT SOURCES

Sources of multiplepoint relations are:

- either images borrowed from sites considered analogous to the one of interest or

- synthetic models generated on the basis of first geological principles

These pictures are called training images. They are absolutely critical to the success of the simulation.

FILTERING

Filters are used to extract essential pattern information, such as average, gradient, and curvature. SGeMS employs six filters for two-dimensional patterns and nine in three-dimensional cases.

Filter weights   applied to   a pattern   generate   a score

Courtesy of A. Journel, Stanford University
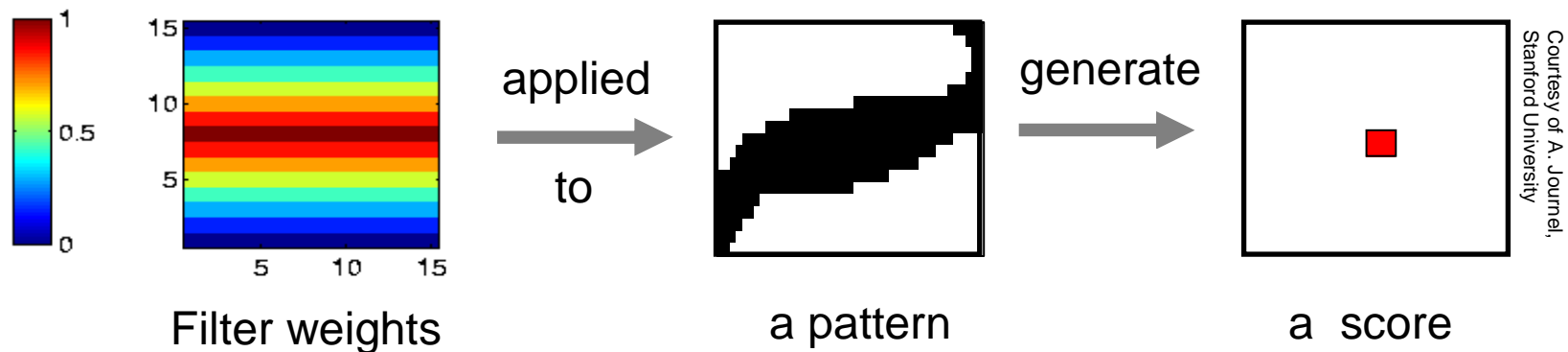
CLUSTER ANALYSIS

The general idea is to group objects in attribute space into clusters as internally homogeneous as possible and as different from the other clusters as possible.

Different types of approaches, distances, and proximity criteria have resulted in several methods. Filter simulation offers two:

- Cross partition, where classes are defined by the quantiles of each filter score. Two patterns belong to the same cluster if they are bounded by the same quantiles for all filters. Results depends on the number of quantiles.

- K-means, which operates by trial and error. Cluster membership evidently varies with the number of clusters.
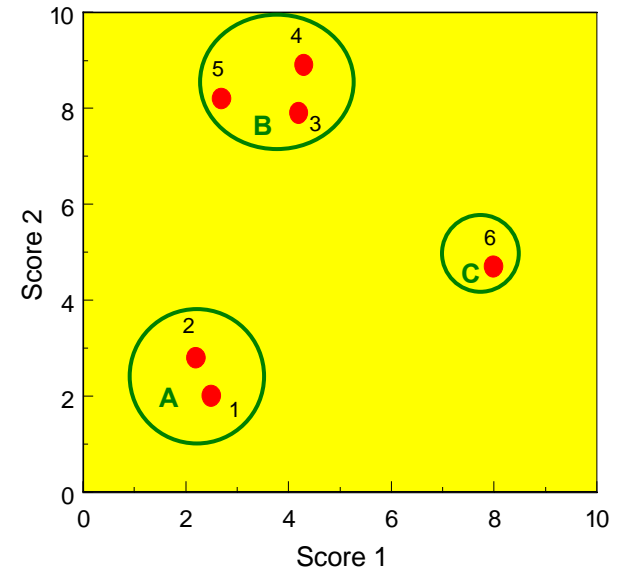
EXAMPLE OF CATEGORICAL CLUSTER

Prototype
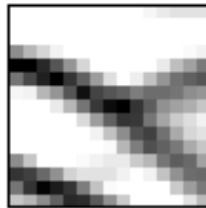
• The prototype here is the average of 56 template patterns from a training image involving two lithologies.

• By the nature of cluster analysis, all the patterns are quite similar.

305
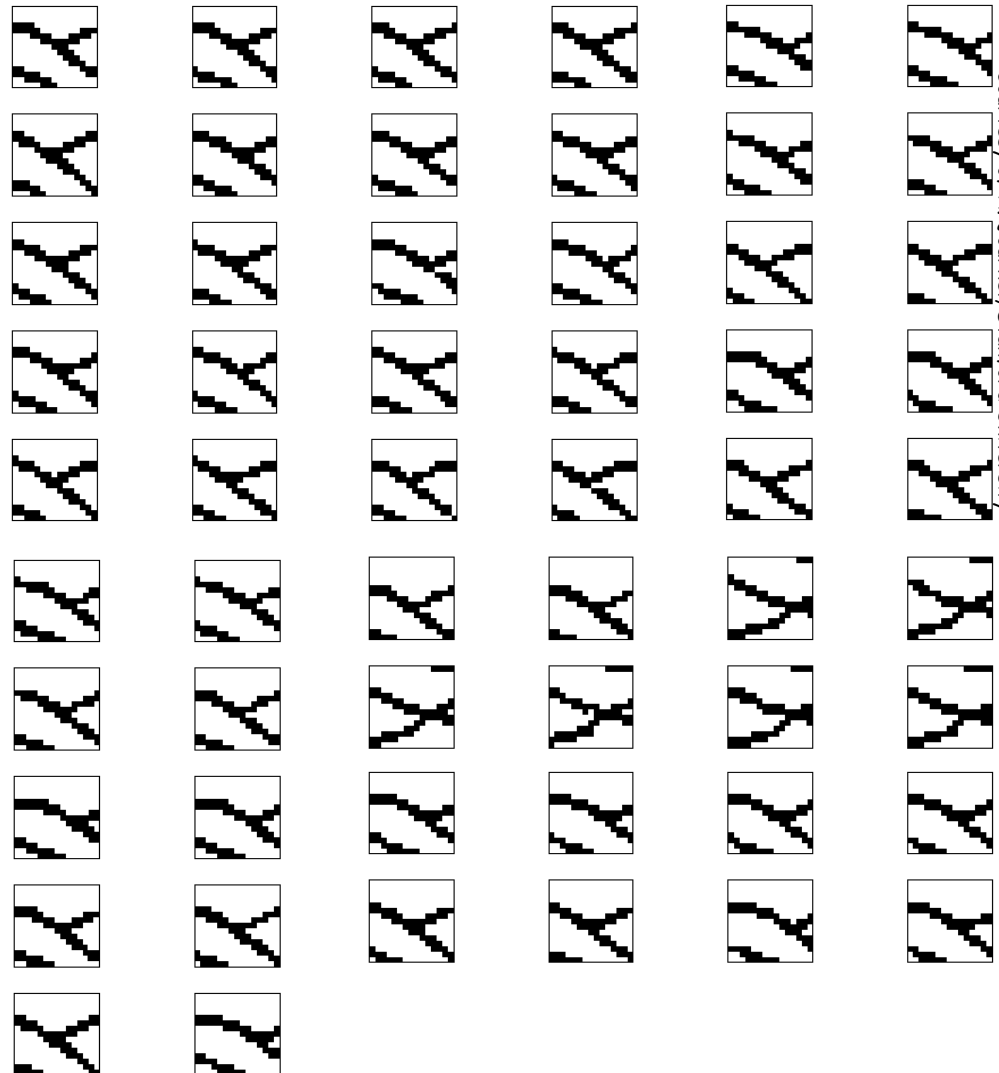
FILTER SIMULATION

Filter simulation is closer to putting together a puzzle than to estimation. The steps to generate a realization are:
1. Define a grid and select a training image at least as large as the study area.
2. Tessellete the training image into small templates.
3. Use different filters to generate scores to characterize template patterns.
4. Use the filtering multivariate information to classify the templates into clusters that group pattern styles.
5. Define a random visitation sequence.
6. Select at random a template from the cluster best matching the conditioning data or select a template at random if there is no conditioning data.
7. Paste the central portion of the template (inner patch) in the realization.
8. Go back to Step 5 until the modeling is completed.

DIAGRAMATIC GENERATION OF AN UNCONDITIONAL REALIZATION

Training image
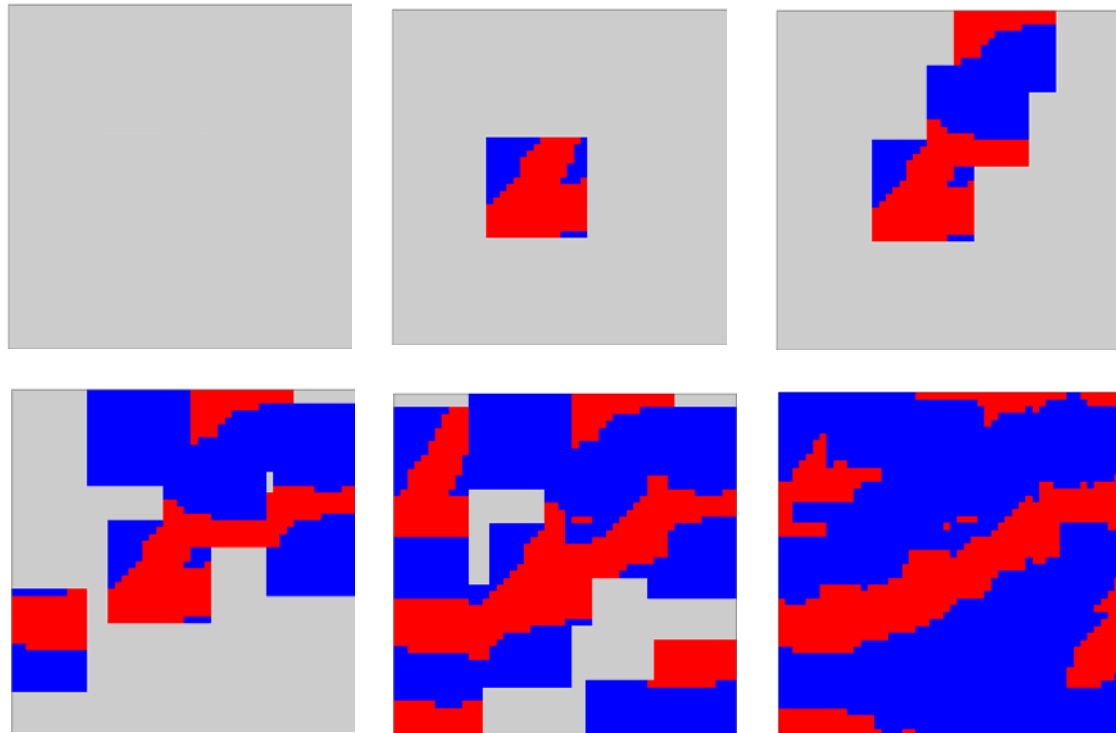
Note that the progression is by patches, not by pixels.

BACK TO THE ELY WEST EXAMPLE

Training image
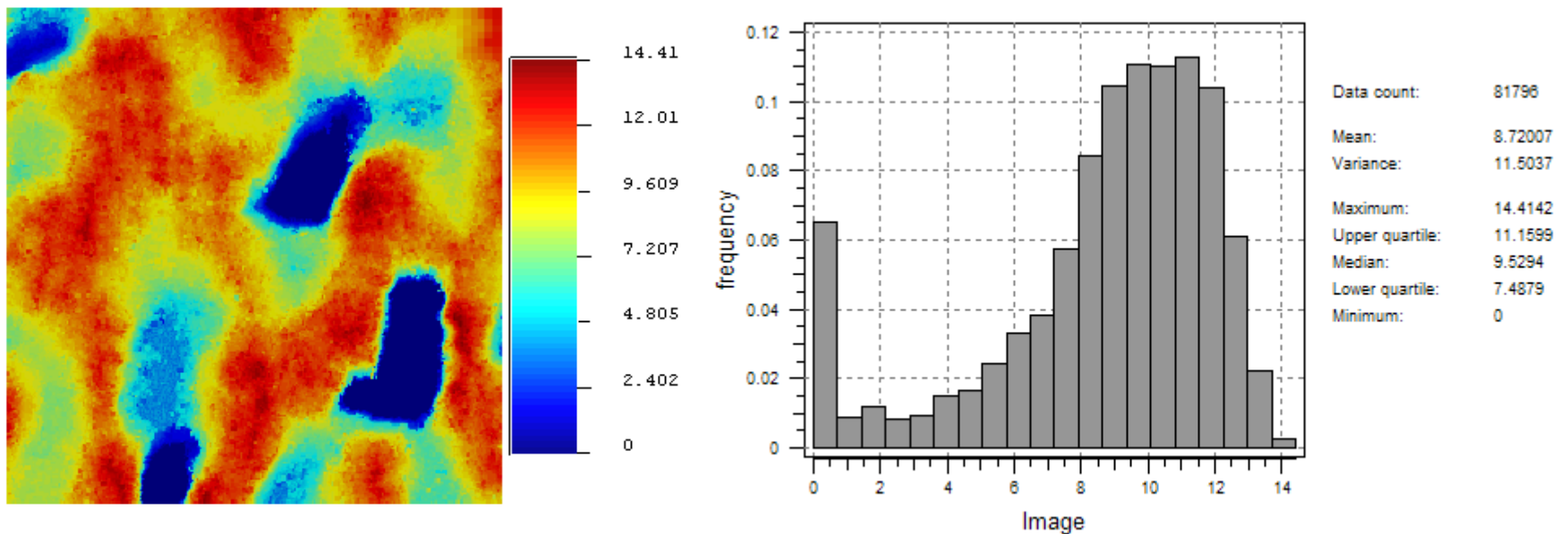
The training image must at least extend over the area of interest.

PAGE 272 EXERCISE ONCE MORE, NOW EMPLOYING FILTER SIMULATION

A. TWO REALIZATIONS FOR ELY WEST

310
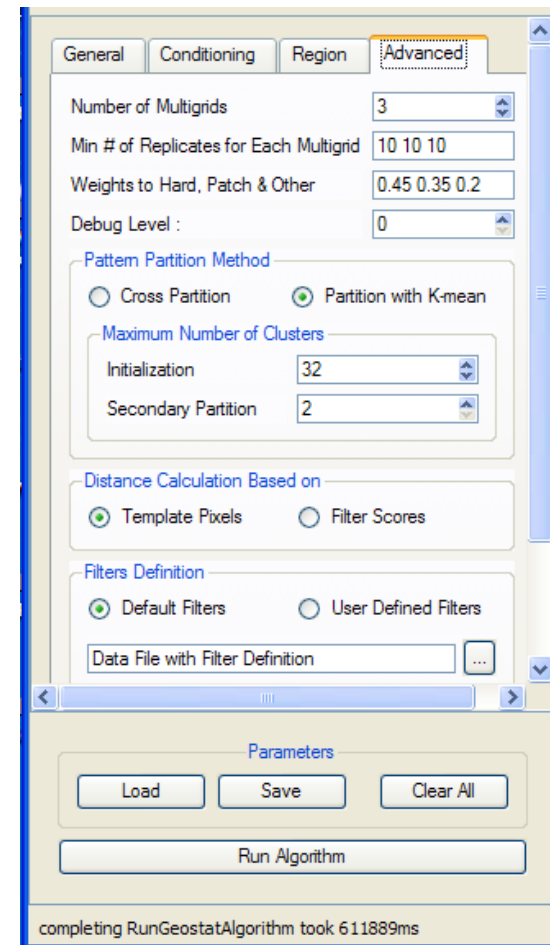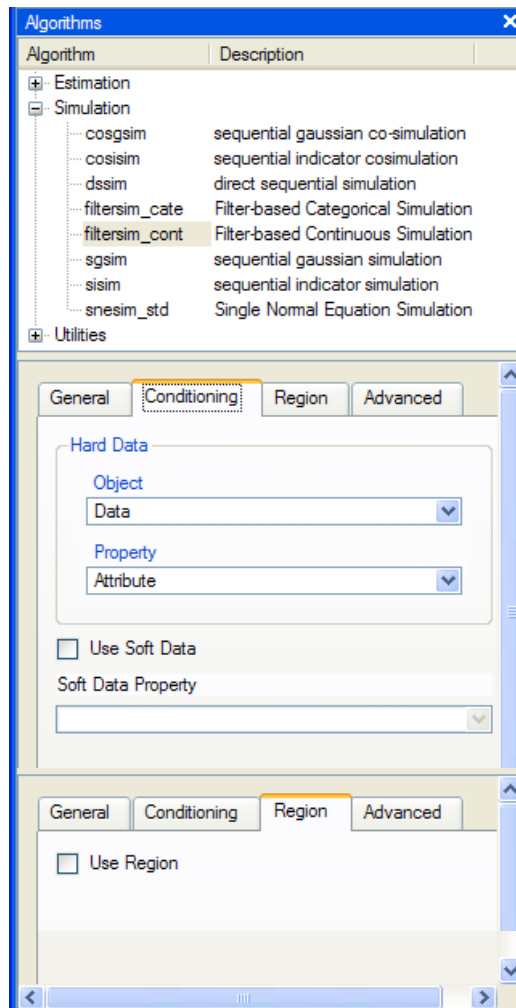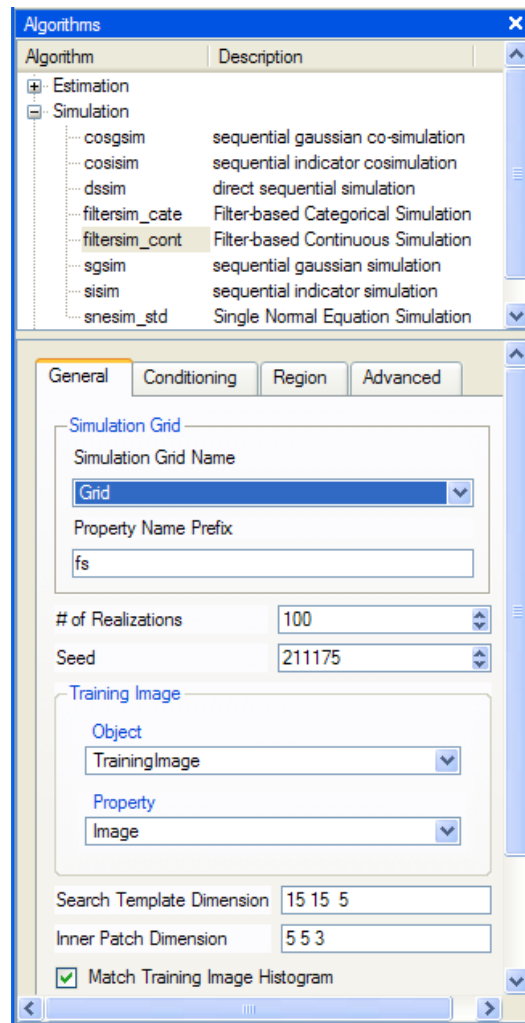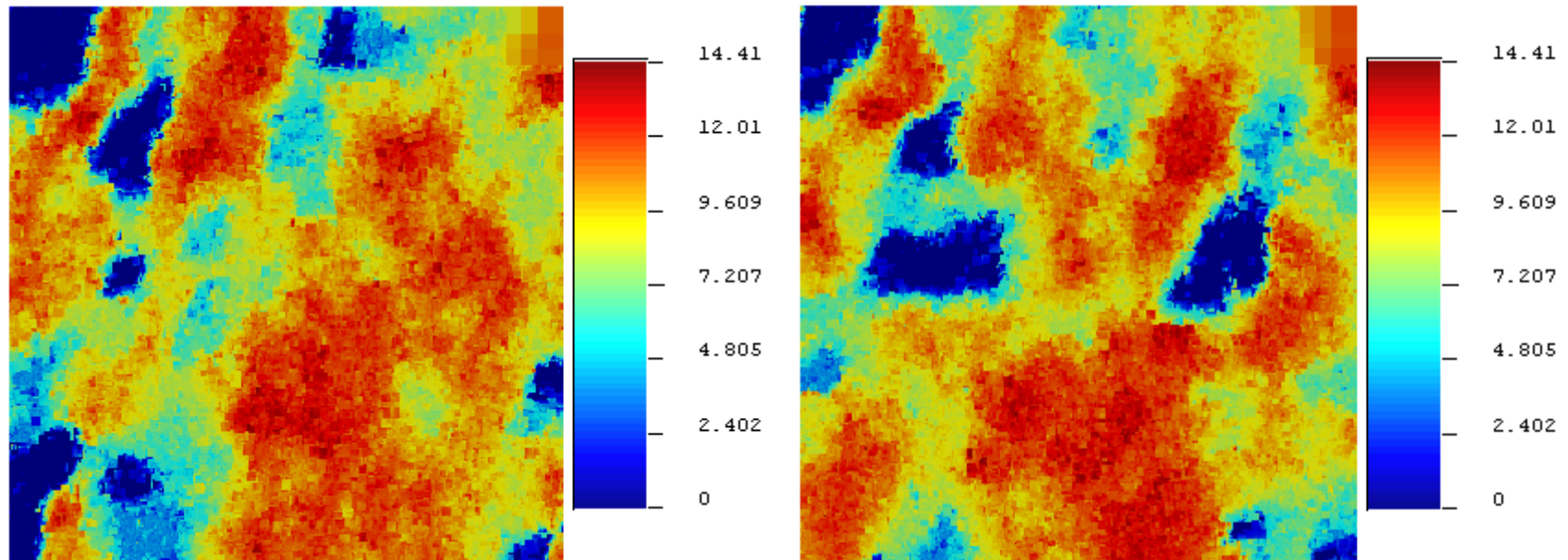
COMPARISON TO EXHAUSTIVE SAMPLE
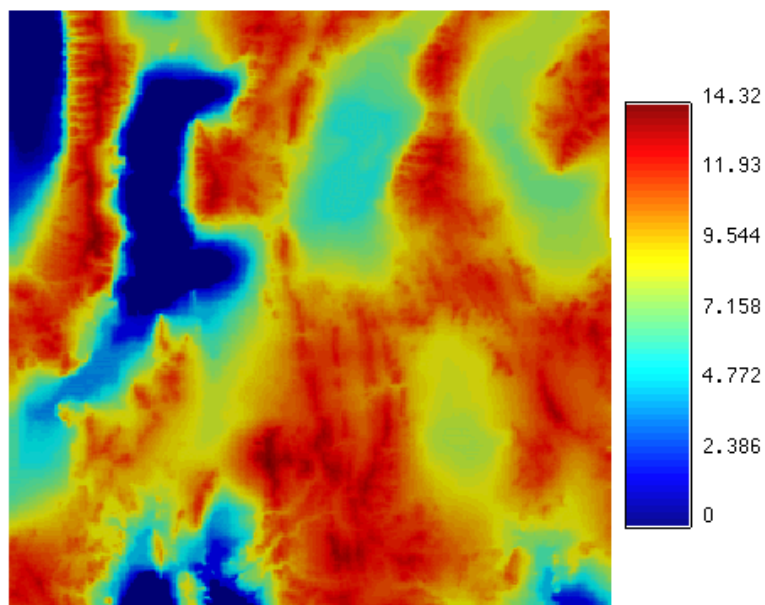
Exhaustive sample
81,796 measurements
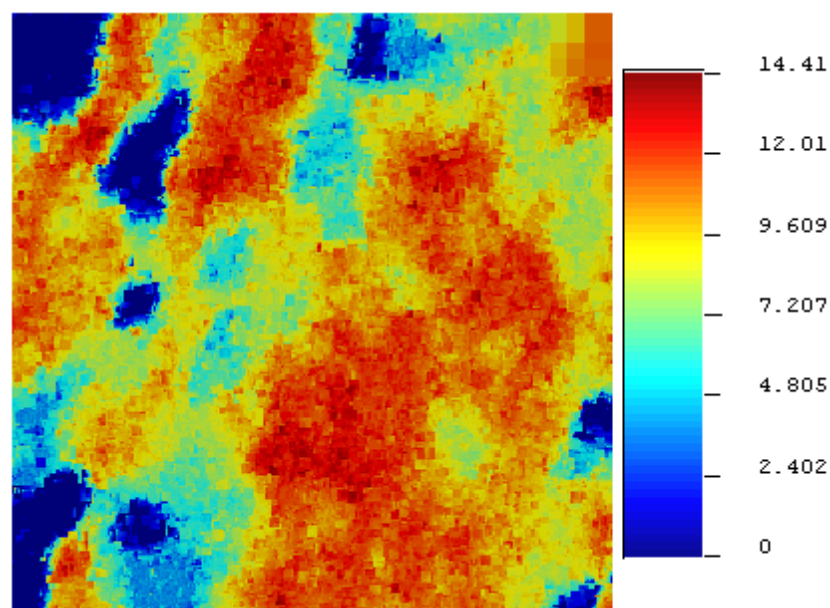
First realization, based on
40 measurements plus
the training image

B. SEMIVARIOGRAMS
FOR FIRST REALIZATION
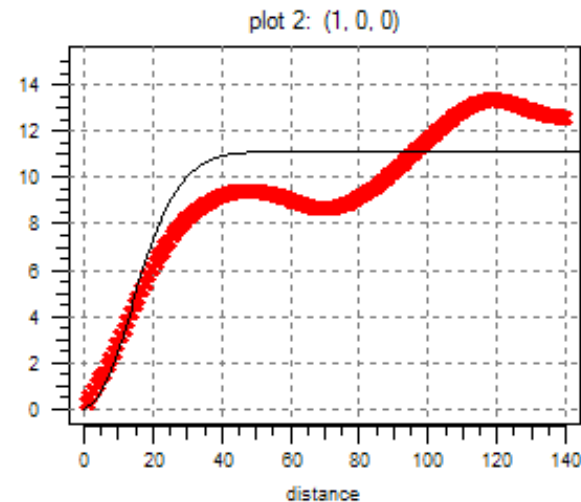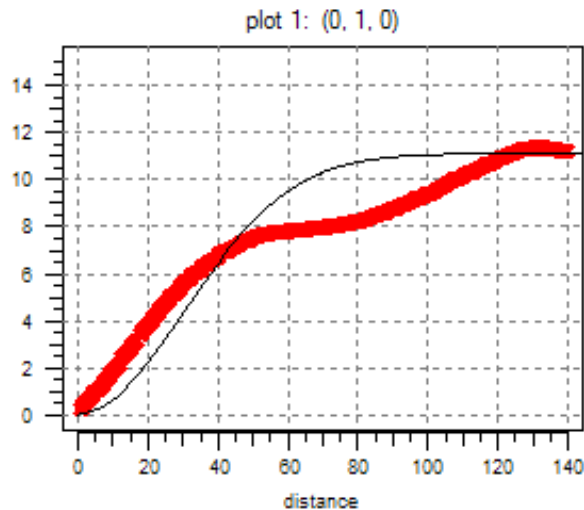
Reproduction of exhaustive semivariogram is not as good as that of simulated annealing, but comparable to the overall approximation by sequential Gaussian simulation.

C. HISTOGRAMS
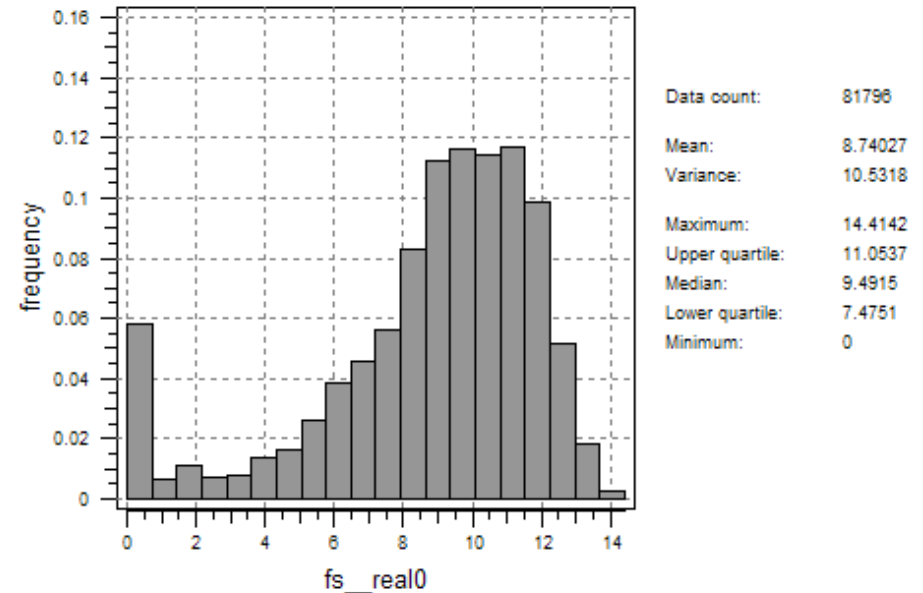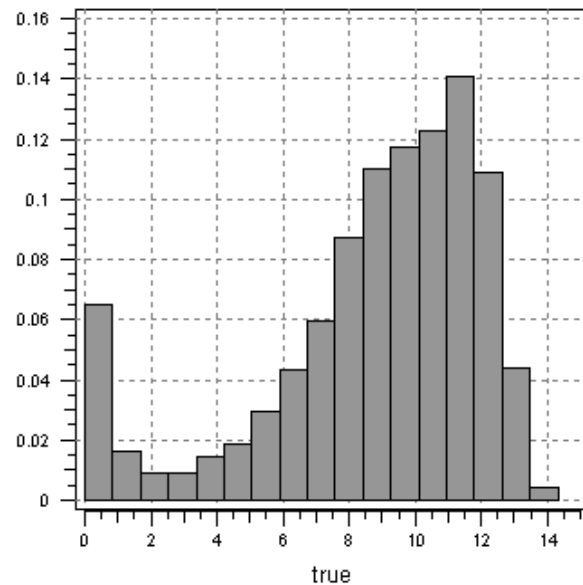
Filter simulation does a superior job approximating the histogram of the exhaustive sample because of the positive influence of the training image.

D. Q-Q PLOT
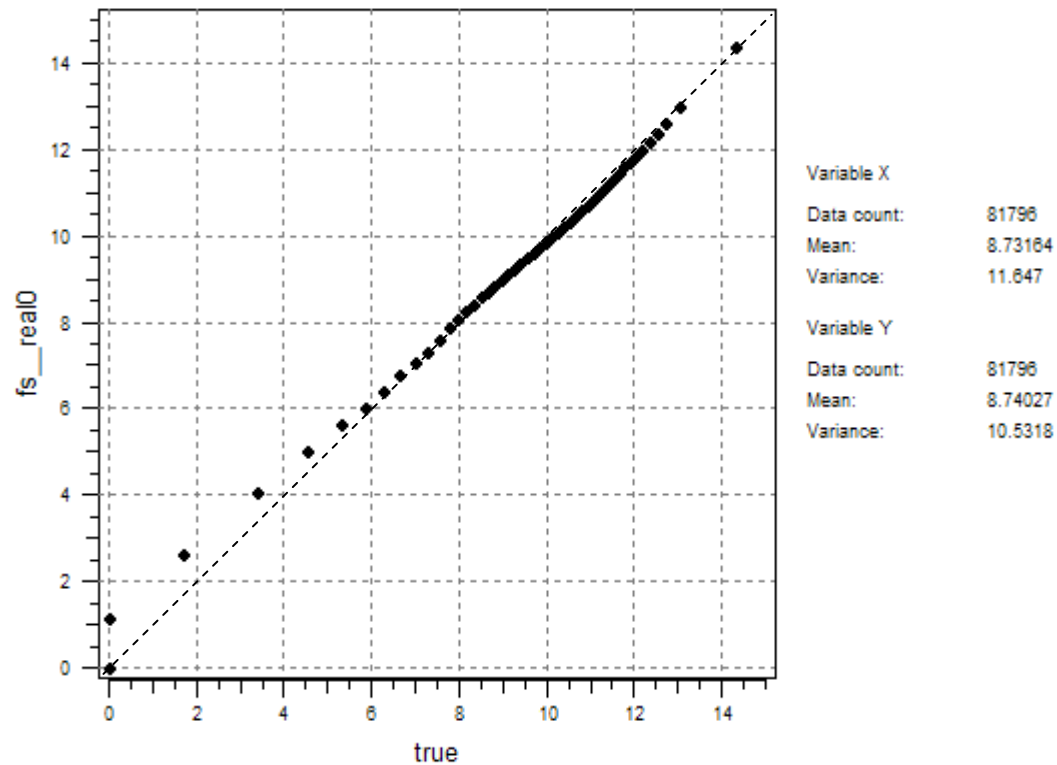
The distribution for the first realization is in close agreement with that of the exhaustive sample.

E. ERRORS IN FIRST REALIZATION ACCORDING TO EXHAUSTIVE SAMPLE
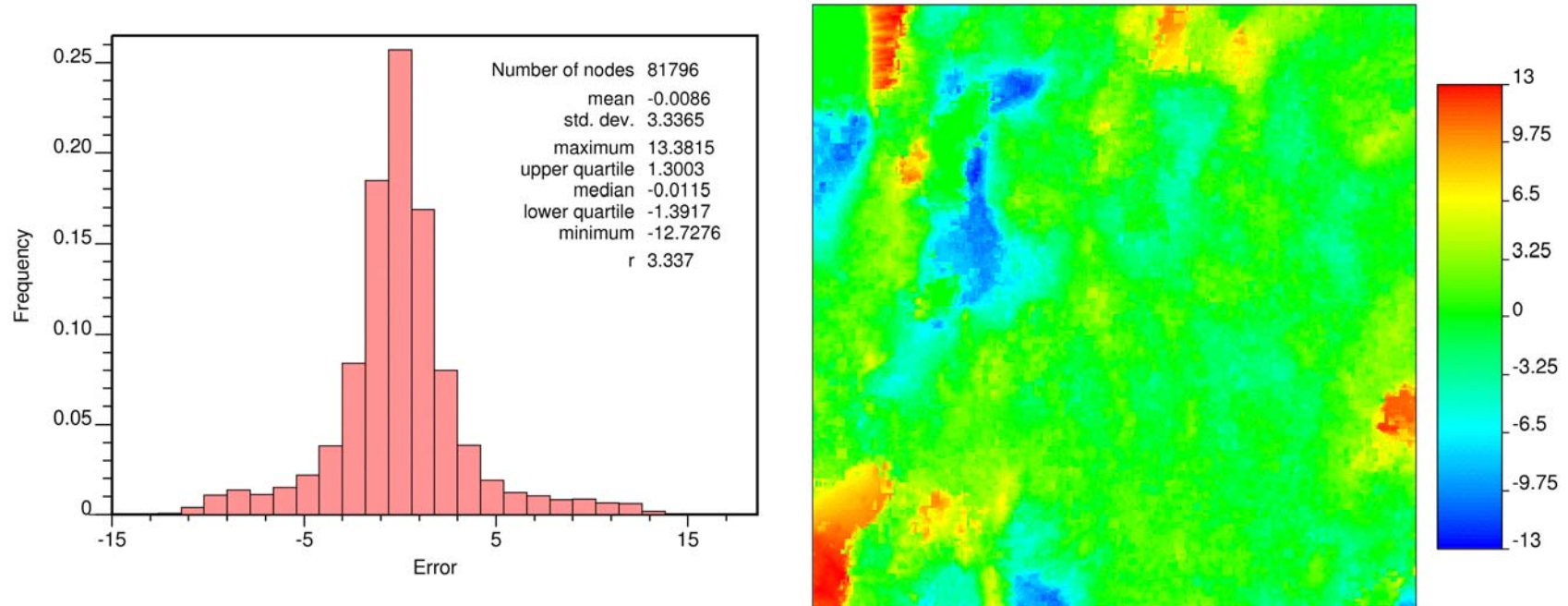
Errors are unbiased and the lowest among the simulation methods considered in this primer.

ERROR SEMIVARIOGRAMS
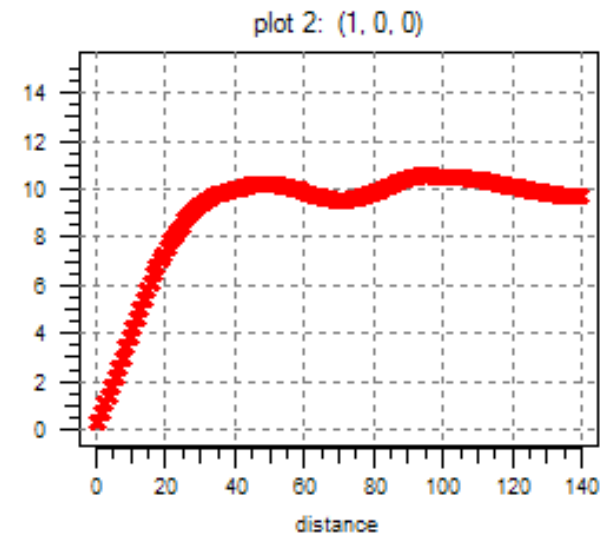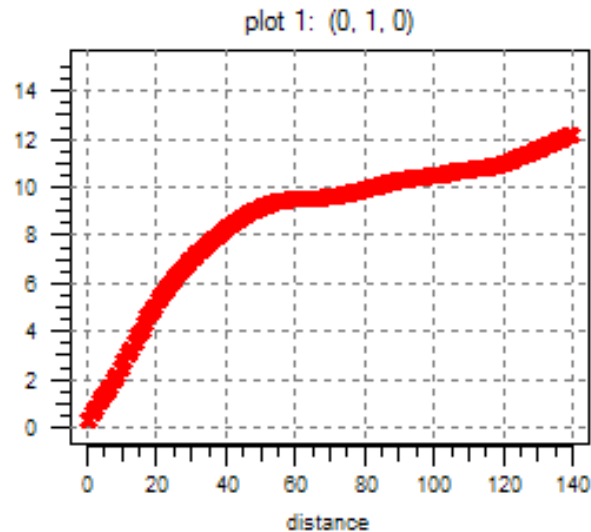
In common with all other methods, errors here are correlated. The semivariogram resembles the one for the attribute being modeled.

316

F. A VIEW OF ERRORS AND NODAL STANDARD DEVIATIONS

Error

Nodal standard deviation
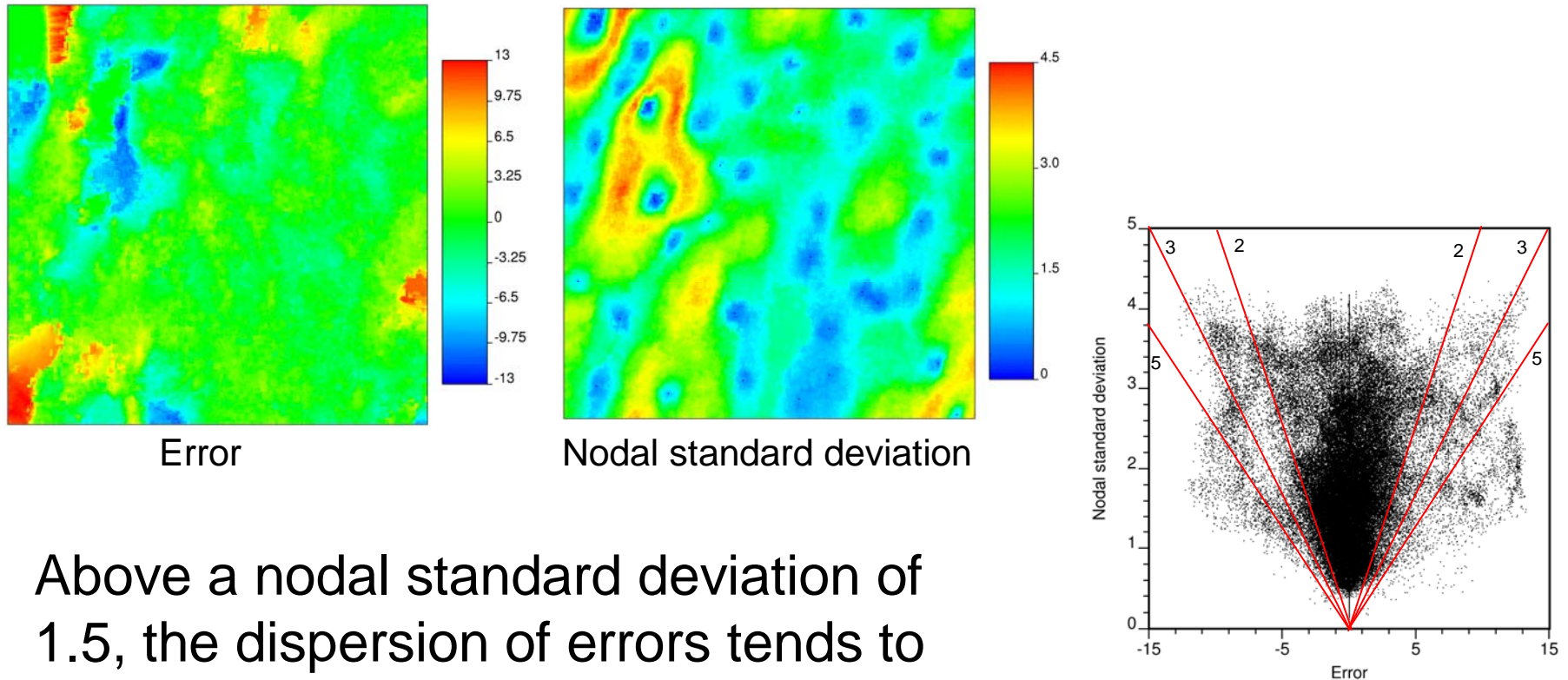
Above a nodal standard deviation of 1.5, the dispersion of errors tends to be the same regardless of the value of the nodal standard deviation.

18. RELIABILITY

318

UNCERTAINTY

From the outset, the quantitative assessment of the uncertainty associated with the resulting models has been an important objective of geostatistics.
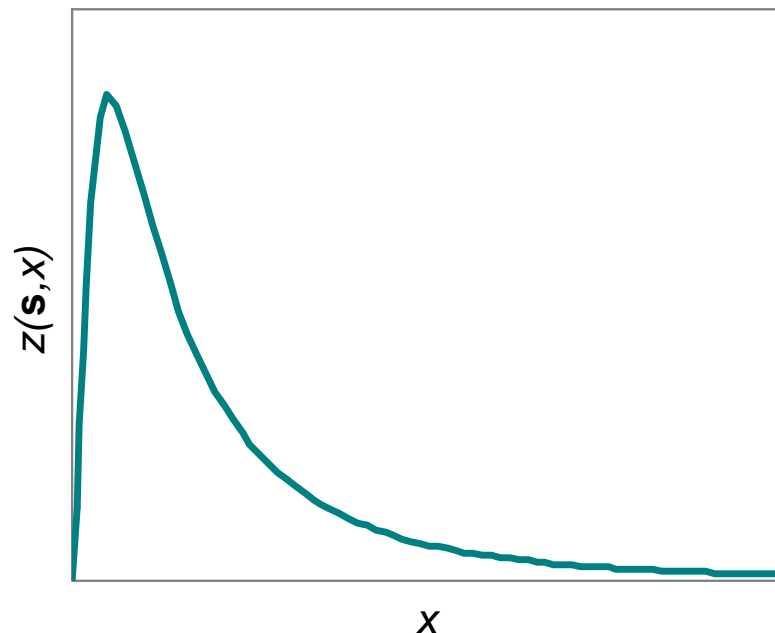
Basic for the analysis is having information about the probability distribution of the errors.

This chapter explores the different ways in which geostatistics uses results from previous methods for uncertainty assessment.

THE RANDOM FUNCTION MODEL

By collecting the values of multiple realizations at the same node, it is possible in practice to model the concept of multiple variates over a sampling domain.

After Christakos and others, 2005.
Reproduced with kind permission of Springer Science and Business Media
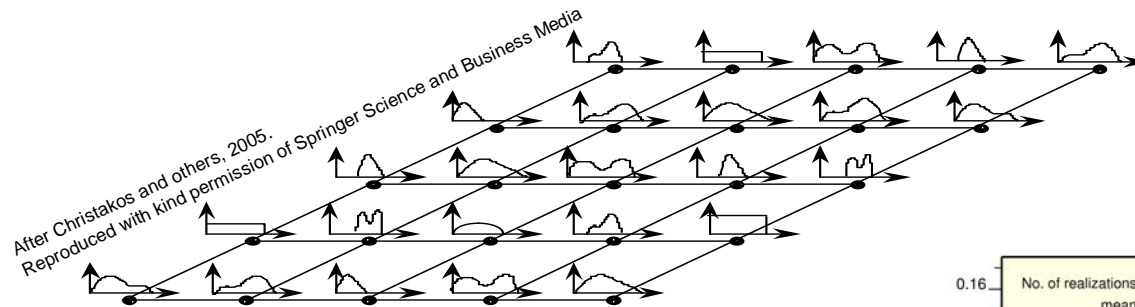
Diagrammatic idea

Actual implementation:
central node, Ely West

No. of realizations 100
mean 10.0622
std. dev. 3.0996
coef. of var 0.3080

maximum 14.0000
upper quartile 12.3647
median 10.8045
lower quartile 8.0930
minimum 0.0000
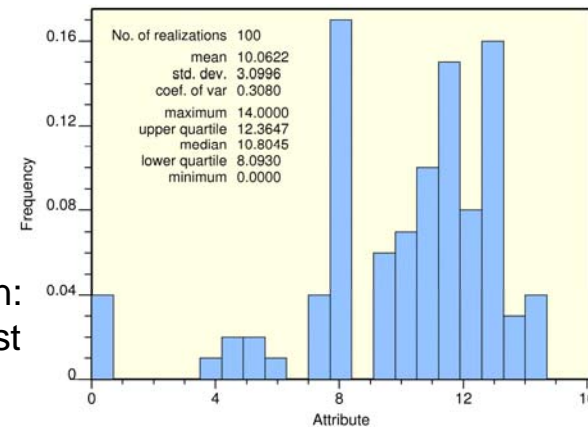
MAPS OF MOMENTS

Although nodal distributions have all the information for assessing uncertainty, customarily there are so many nodes that their individual inspection is impractical, creating, once more, a need to prepare summaries.

   The easiest solution is to map some consistent summary statistic for the nodal distributions. The most common options are:

A. Nodal mean, also known as E-type
B. Nodal standard deviation

   Both options share the advantage of reducing all information to one map, no matter how many the realizations.

A. NODAL MEAN MAP

- E-type maps tend to remind us of kriging maps, sharing their smooth appearance.

- As in kriging, smoothing disqualifies E-type maps from being realistic models.

- In the unlikely event that an E-type map is a good model, it is always less demanding to prepare it by using kriging.

  The link between E-type maps and kriging is an interesting fact to remember, but E-type maps lack practical interest.

Ordinary kriging estimation

Number of nodes 14700
correlation 0.988
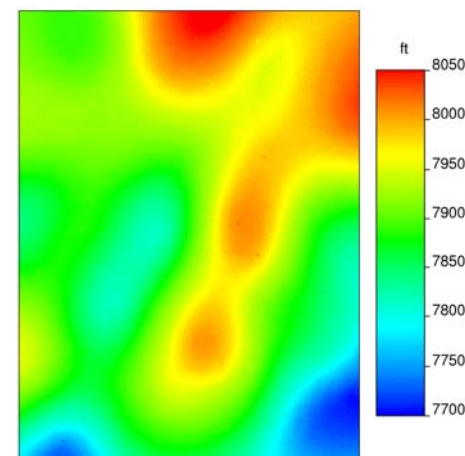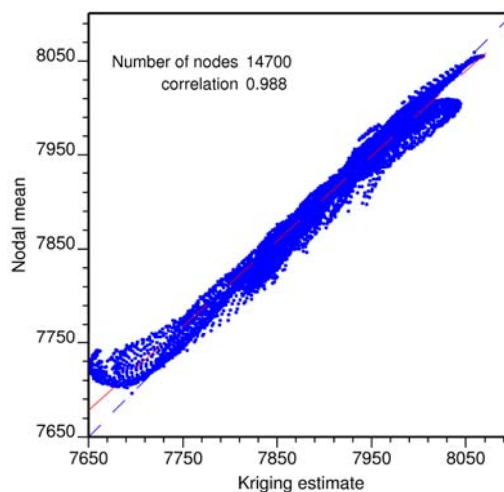
Sequential Gaussian simulation

B. NODAL STANDARD DEVIATION MAPS

Ordinary kriging estimation

Sequential Gaussian simulation
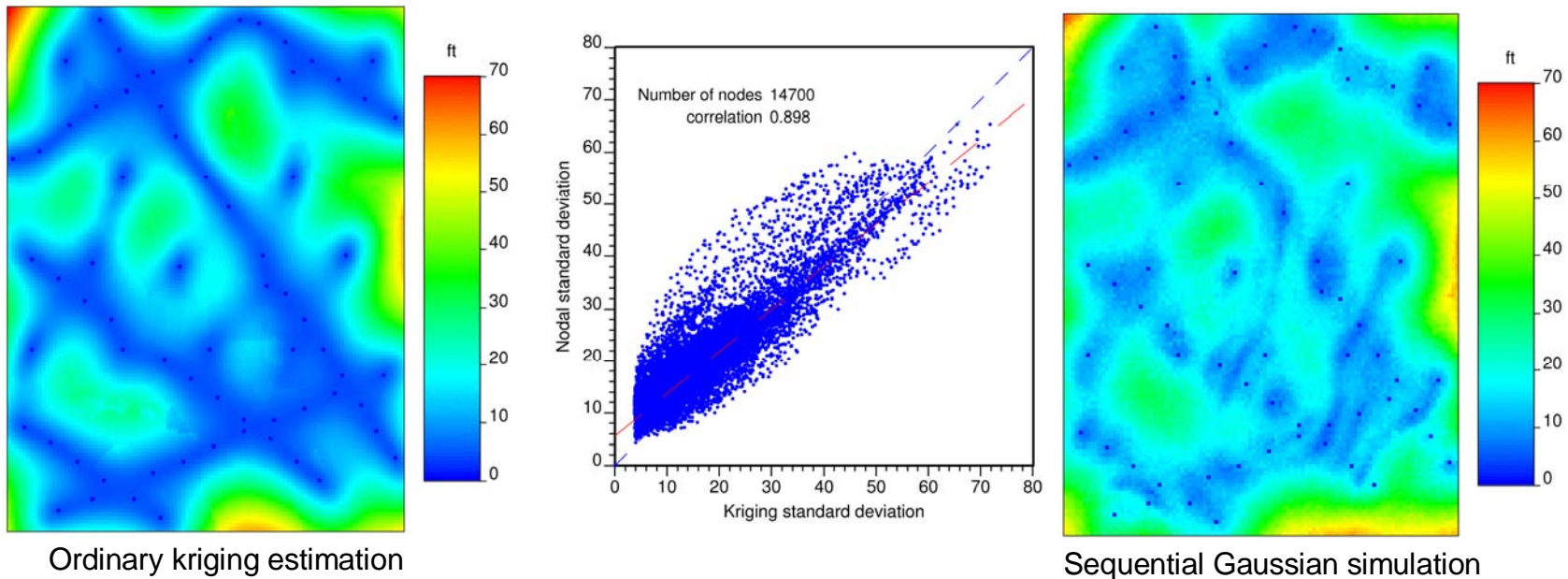
- The resemblance between standard deviations from simulation and kriging can be good, but never as good as the one for the means.
- Although nodal standard deviation provides a semi-quantitative feeling about uncertainty, other summary statistics are more efficient for summarizing the information contained in the realizations.

NORMALITY

Kriging can be used in reliability estimation, but it is necessary to assume the form of the nodal distributions. The common practice is to assume normality, with a mean equal to the estimate, z^*, and a standard deviation equal to the kriging standard error, σ^*.
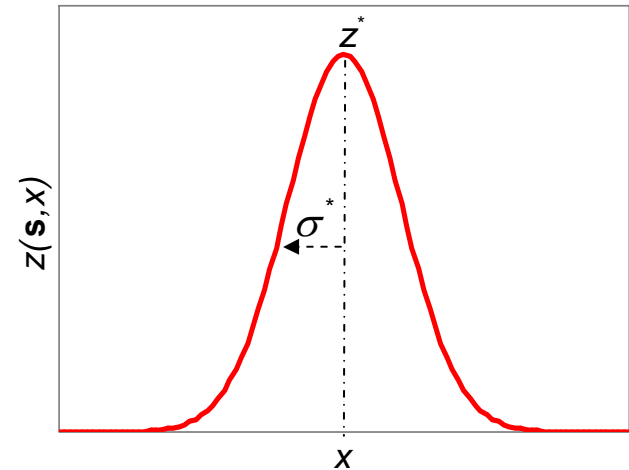
When the errors indeed follow a multivariate normal distribution reasonably closely, results can be quite reliable, as we will see for the case of the UNCF sample.

Whereas kriging may not be great at characterizing the true attribute, it can do a much better job, comparable to that of simulation, when it comes to assessing confidence intervals.

BEST DISPLAYS FOR SUMMARIZING UNCERTAINTY

A. The probability distribution for some global attribute after reducing each realization to one single number, which can be of the same nature or different from the sampled attribute.

B. The kth percentile map displaying the threshold for which there is a probability $0 \leq k/100 \leq 1$ that the true value is less than or equal to the threshold. The units of this type of map are the same as those of the attribute.

C. A probability map showing the chances that the true value be equal to or less than a given threshold.

D. Interval probability map displaying fluctuations in the chances that the true value be between a pair of fixed values.

A. SUMMARY DISTRIBUTIONS ASSESSING GAS POTENTIAL IN GREATER NATURAL BUTTES, UTAH
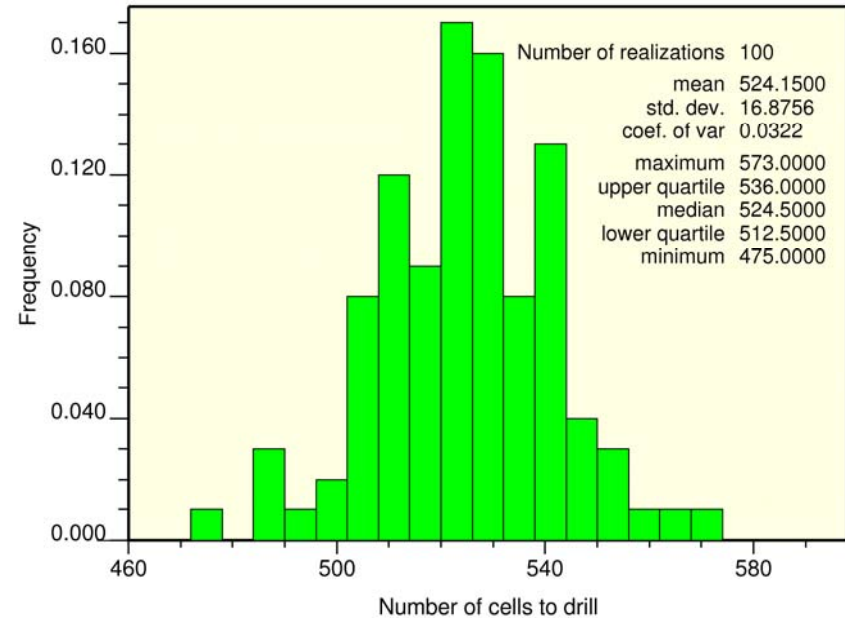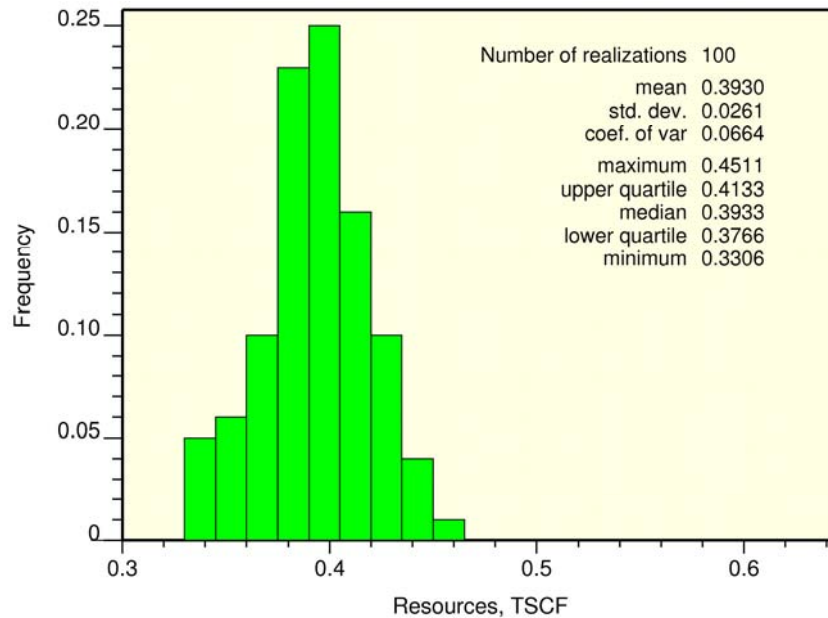
B. BACK TO THE UNCF EXAMPLE.
10th PERCENTILE MAP

$$\text{Prob}[Z(\mathbf{s}) \leq node\_value(\mathbf{s})] = 0.10$$
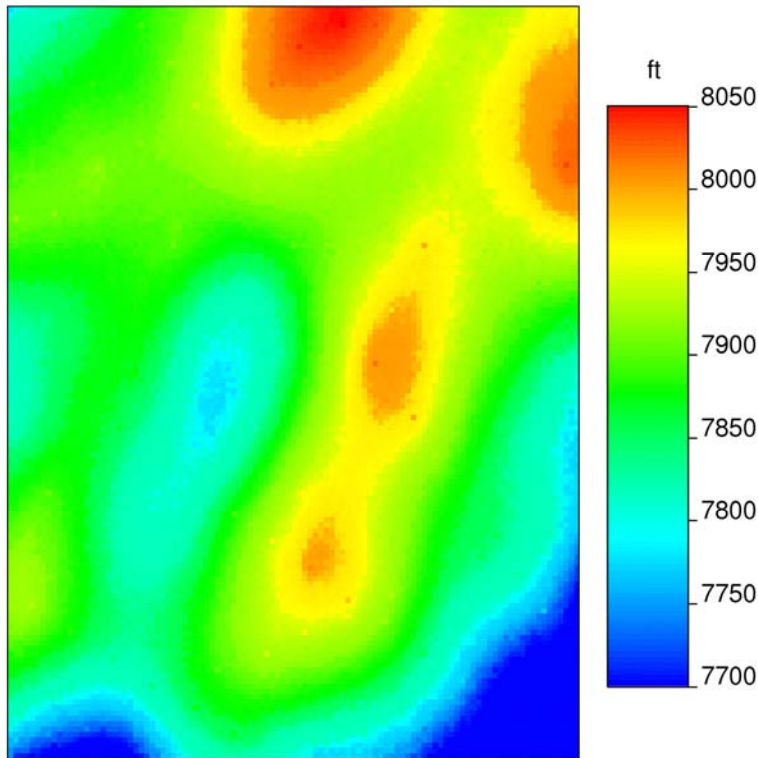
Sequential Gaussian simulation
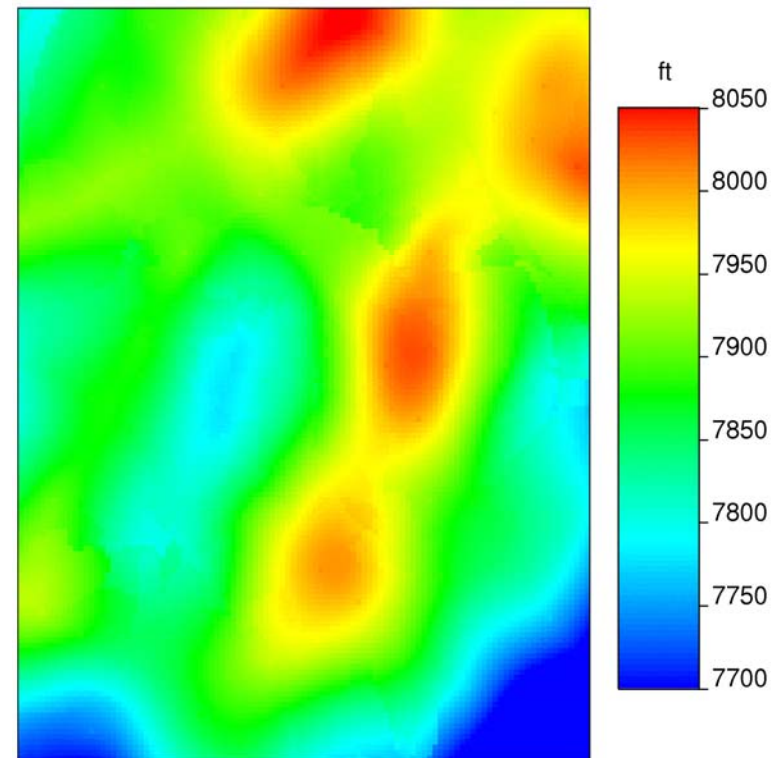
Ordinary kriging estimation

C. PROBABILITY MAP FOR TWO THRESHOLDS USING SEQUENTIAL GAUSSIAN SIMULATION
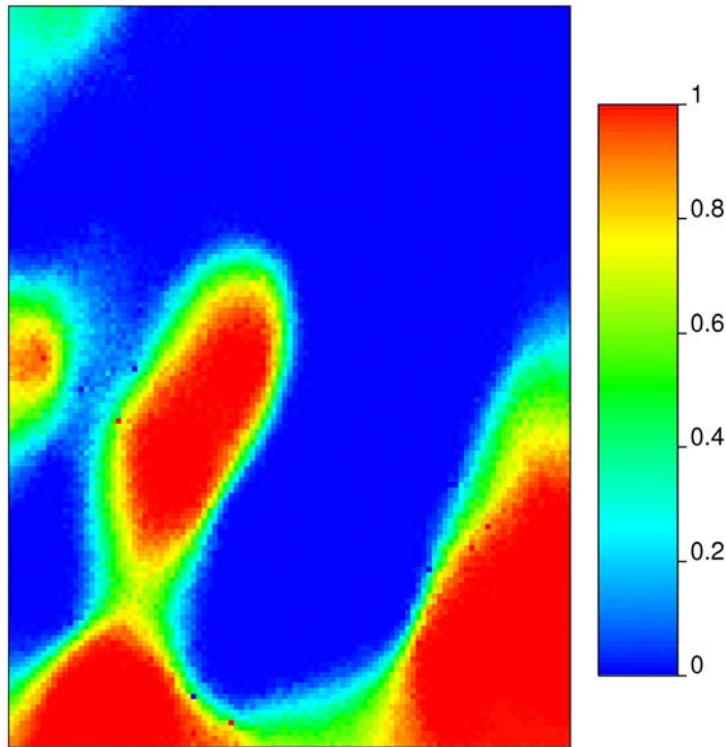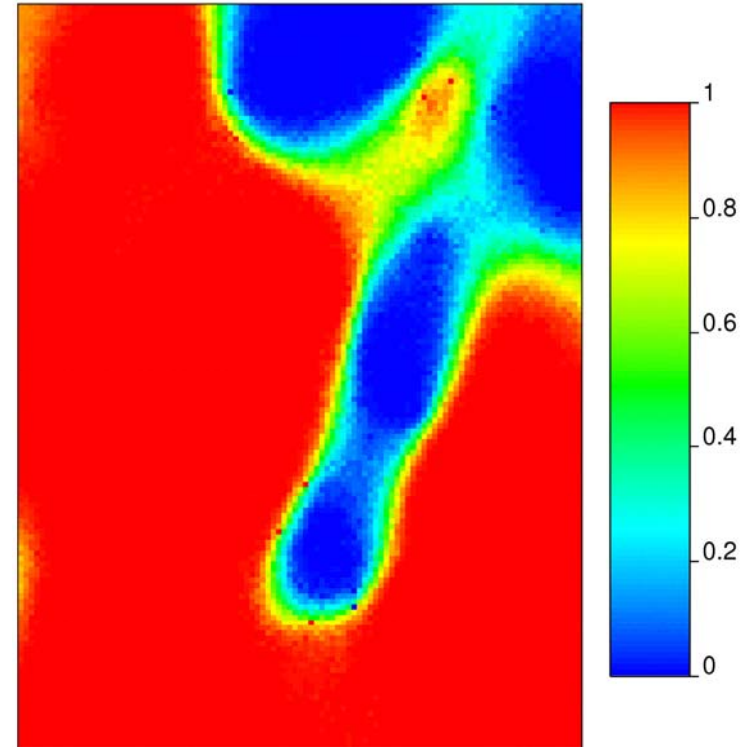
$$\text{Prob}\big[Z(\mathbf{s}) \leq 7872\big] \qquad\qquad \text{Prob}\big[Z(\mathbf{s}) \leq 7966\big]$$

PROBABILITY MAP FOR TWO THRESHOLDS UNDER NORMALITY
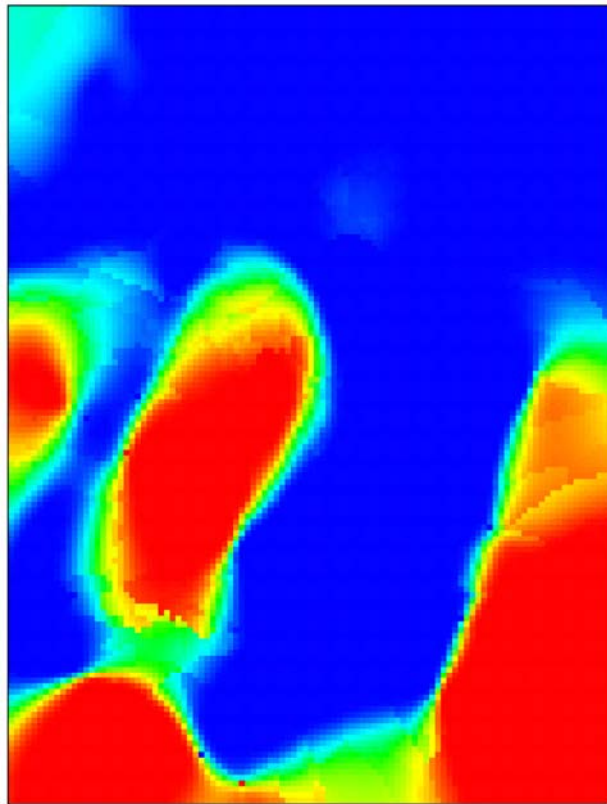
$\text{Prob}[Z(\mathbf{s}) \le 7872]$
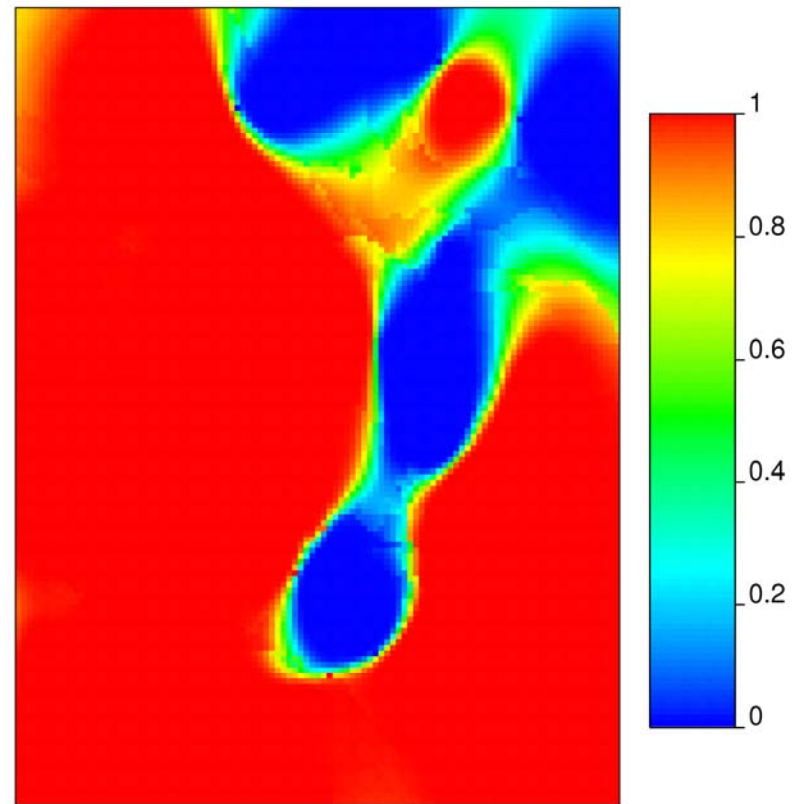
$\text{Prob}[Z(\mathbf{s}) \le 7966]$

D. INTERVAL PROBABILITY

$$\text{Prob}\left[7872 \le Z(\mathbf{s}) \le 7966\right]$$
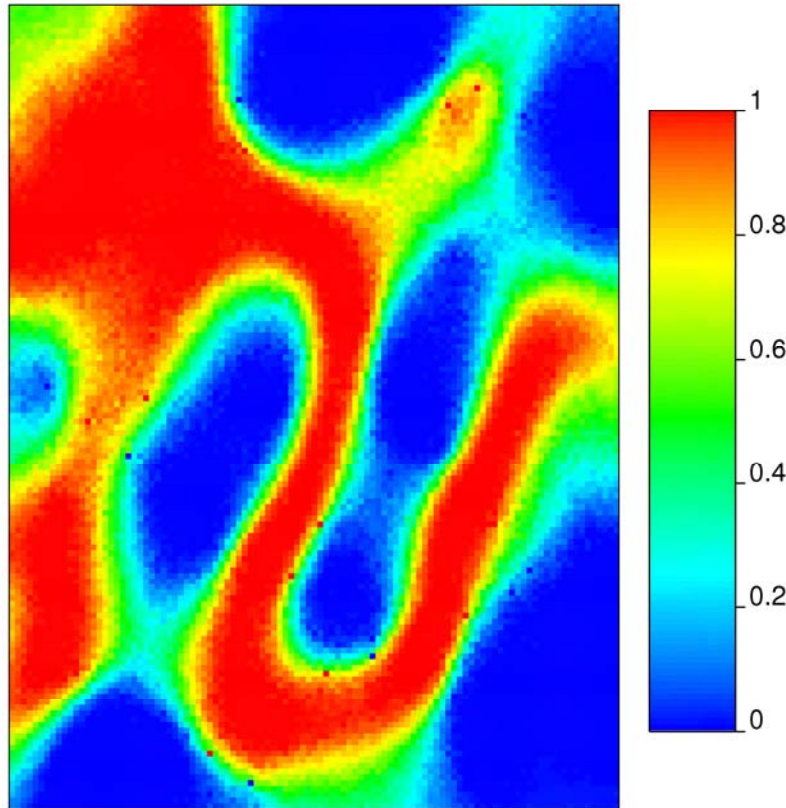
Sequential Gaussian simulation
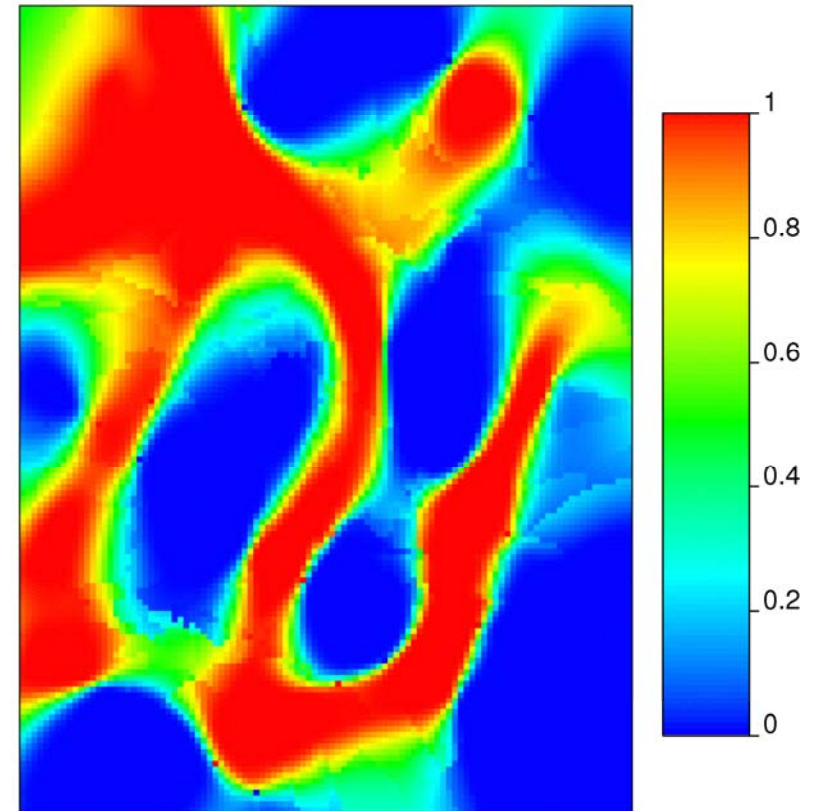
Ordinary kriging estimation

BACK TO ELY: WHERE ARE TRUE VALUES LANDING AT THE NODAL DISTRIBUTIONS?
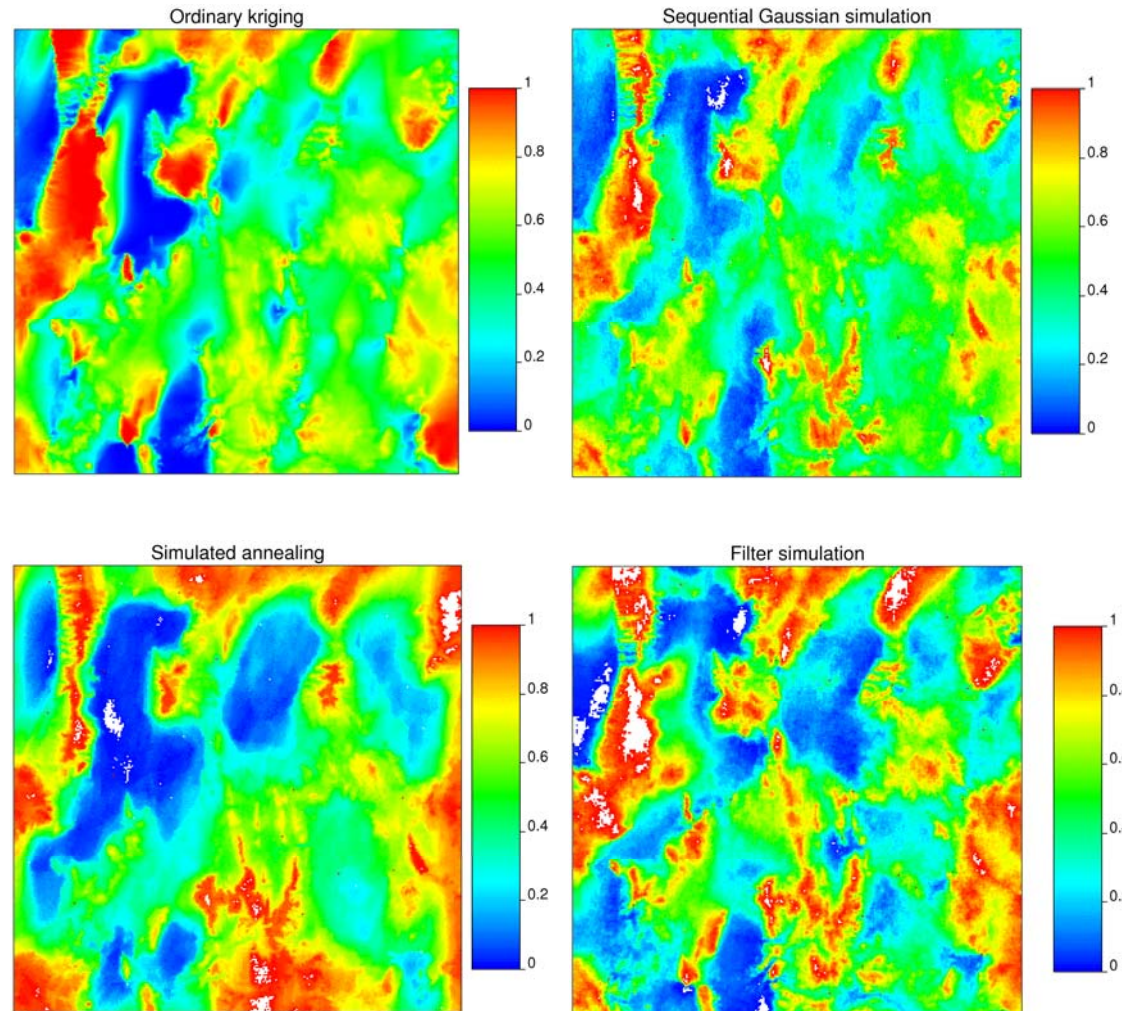
Each node shows the cumulative probability below the true value.

Blanks denote that the true value is outside the nodal distribution, certainly not a good situation.

ACCURACY OF CONFIDENCE INTERVALS
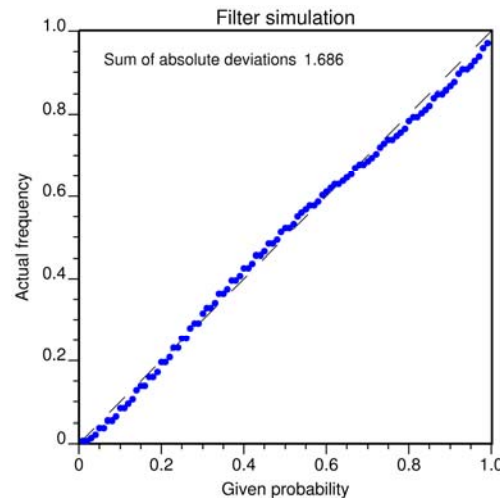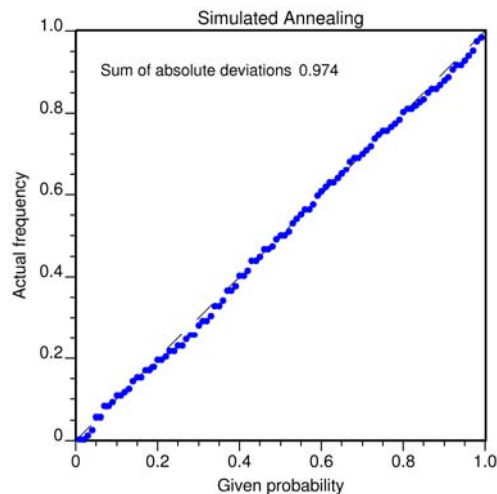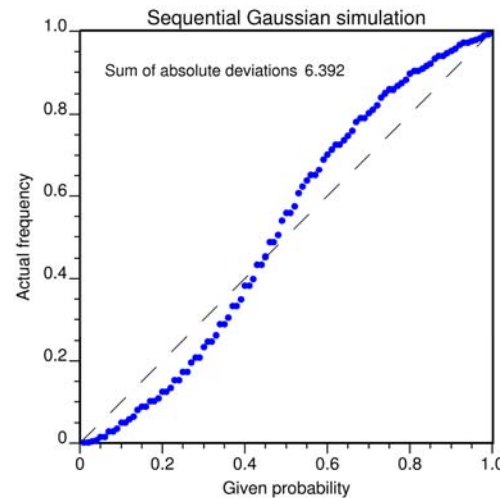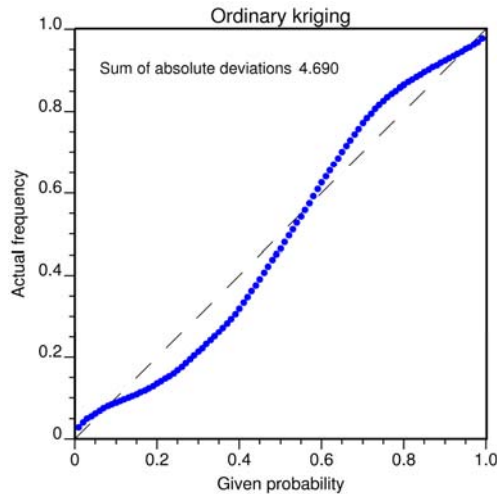
Using the nodal distributions, for a given probability, the scatterplots show how frequently the true values are indeed below the quantile associated with the given probability.

Note that these plots are global assessments comprising all nodes for all realizations.

19. NAVIGATION CHARTS

KRIGING OR SIMULATION?

Although there are no clear excluding rules, common criteria to prefer one approach over the other are:

- Perform kriging if the paramount criterion is individual minimization of prediction error or if the main interest is an exploratory mapping of a poorly known attribute through a deliberately smooth, thus simplified, version of reality.

- Go for simulation if it is not acceptable having to assume normality of the errors for reliability assessment or if correct modeling of spatial continuity is more important than local accuracy.

Always remember that, unless you are in a terrible hurry, nothing prevents trying both kriging and simulation, as done with Ely data, and judge according to the results.

RANKING OF FOUR METHODS
IN THE SPECIAL CASE OF ELY WEST

Method	Distribution reproduction	Semivariogram reproduction	Texture	Transition	Error	Confidence interval	Total
1. Filter simulation	1	2	1	1	2	2	9
2. Sim. annealing	2	1	3	2	3	1	12
3. Seq. Gauss. sim.	2	2	2	4	3	4	17
4. Ordinary kiriging	4	4	4	2	1	3	18

Exhaustive sample
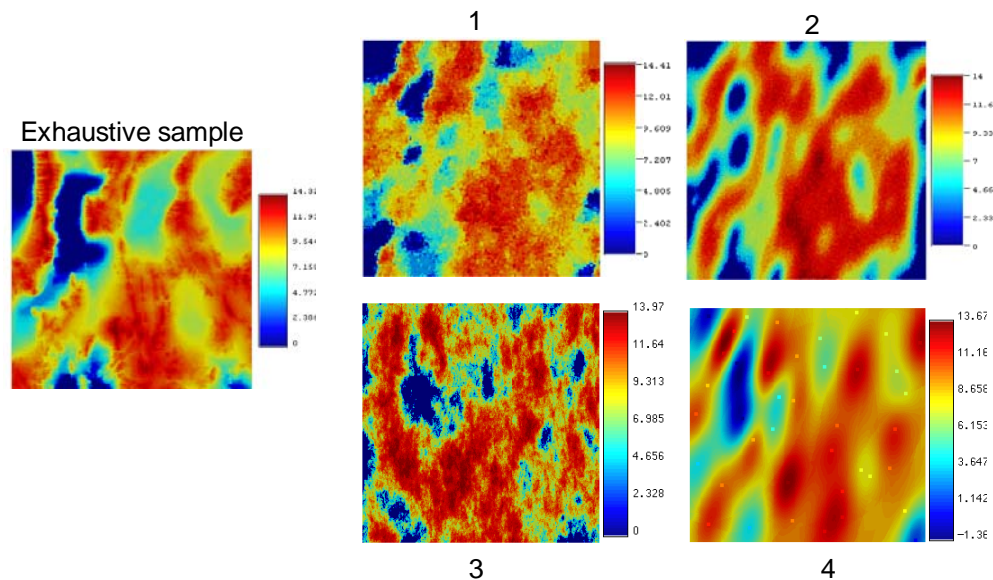
WHAT TYPE OF KRIGING?

Selection of the type of point kriging is ruled by the knowledge of and type of mean (trend).

- Use simple kriging if there is no trend and the population mean is known, a rare situation.
- Run ordinary kriging if the mean is unknown and the trend is approximately constant within the search neighborhood (page 185). This is the predominant case. Ordinary kriging is actually a special name for universal kriging for a polynomial of degree 0.
- Otherwise, employ universal kriging proper of degree 1 or 2, whose main feature is extrapolation, which is still poor.

Use the same criteria to decide the type of block kriging and cokriging.

As the complexity of the trend model goes up from simple to universal kriging, so does the estimation variance.

As for the use of semivariogram or covariance, the choice is immaterial because they are equivalent.

WHAT KIND OF SIMULATION?

- Apply filter simulation if you already have a good idea about the spatial variation of your attribute, good enough to be able to provide a training image. Quality of results are crucially dependent on the quality of the training image.

- In terms of usage, popularity of the methods is divided among the other two types covered here. Go for simulated annealing if you want to stay away from any form of normality assumption and the grid size is less than 100,000 nodes.

- Otherwise, try sequential Gaussian simulation.

ACKNOWLEDGMENTS

SELECTED BIBLIOGRAPHY

Caers, J., 2005, Petroleum geostatistics: Richardson, Tex., Society of Petroleum Engineers, 88 p.

Chilès, J.-P. and Delfiner, P., 1999, Geostatistics: modeling spatial uncertainty: New York, Wiley-Interscience, 695 p.

Christakos, G., Olea, R.A., Serre, M.L., Yu, H.-L., and Wang, L.-L., 2005, Interdisciplinary public health reasoning and epidemiologic modelling: Berlin, Springer, 319 p.

Deutsch, C.V. and Journel, A.G., 1992, GSLIB: geostatistical software library and user's guide: New York, Oxford University Press, 340 p. and 2 diskettes; 2nd ed., 1998, 369 p. and CD-ROM.

Goovaerts, P., 1997, Geostatistics for natural resources evaluation: New York, Oxford University Press, 483 p

Isaaks, E.H. and Srivastava, R.M., 1989, An introduction to applied geostatistics: New York, Oxford University Press, 561 p.

Moore, D.S. and Notz, W.I., 2006, Statistics: concepts and controversies (6th ed.): New York, Freeman, 561 p.

Olea, R.A., ed., 1991, Geostatistical glossary and multilingual dictionary: New York, Oxford University Press, 177 p.

Olea, R.A., 1999, Geostatistics for engineers and earth scientists: Norwell, Mass., Kluwer Academic Publishes, 313 p.

Remy, N., Boucher, A., and Wu, J., 2009, Applied geostatistics with SGeMS: a user's guide: New York, Cambridge University Press, 284 p. and CD-ROM.

INDEX